PROCEEDINGS OF THE CONFERENCE ON DOCUMENTATION AND INFORMATION RETRIEVAL IN HUMAN SENSORY PROCESSES

MARCH 1972

HV1721
K413

EX LIBRIS

M.C. MIGEL LIBRARY
AMERICAN PRINTING
HOUSE FOR THE BLIND

HV1721
K413

PROCEEDINGS OF THE CONFERENCE ON DOCUMENTATION AND INFORMATION RETRIEVAL IN HUMAN SENSORY PROCESSES

Sponsored by
American Foundation for the Blind
University of Pittsburgh

Allen Kent, Godfrey D. Stevens, Leslie L. Clark, Editors

American Foundation for the Blind 1972

A NOTE FROM THE PUBLISHER

The "information explosion" is affecting every profession and field of endeavor. It is particularly acute in fields such as work for the blind since so many disciplines are involved. The American Foundation for the Blind hopes, through the publication of this PROCEEDINGS OF THE CONFERENCE ON DOCUMENTATION AND INFORMATION RETRIEVAL IN HUMAN SENSORY PROCESSES, to provide concerned persons with a better understanding of the current state of documentation as it relates to blindness, and to encourage discussion about and the implementation of solutions to the problem.

M. Robert Barnett
Executive Director
American Foundation for the Blind

These *Proceedings* are a publication of the

American Foundation for the Blind
15 West 16th Street
New York, New York 10011

Leslie L. Clark, Editor
International Research
Information Service

CONTENTS

FOREWORD v

PREFACE vii

NETWORKS, TECHNOLOGY, AND OVERALL FUTURES
Jordan Baruch 1

ERIC - A TOOL FOR RESEARCHERS IN SENSORY PROCESSES
Harvey Marron 5

RELATIONSHIPS OF LARGE SYSTEMS TO SPECIALIZED INFORMATION CENTERS
Joseph Leiter 13

SWITCHING AND TRANSFER PROBLEMS
George Howick 21

A SURVEY OF INFORMATION SYSTEMS ORIENTED TO BLINDNESS
Samuel T. Price 23

A DETAILED DESCRIPTION OF THE IRIS DOCUMENT CENTER
James W. Linsner 27

THE VISION INFORMATION CENTER
Robert Reinecke 31

THE CENTRALIZATION, DECENTRALIZATION, SPECIALIZATION ISSUE
Allen Kent 35

CONCERNS OF COMPATIBILITY AND ANALYSIS OF RESEARCH:
ARTICULATION OF INFORMATION SYSTEMS
Ted Ward 41

RESPONSE OF A SYSTEM OPERATOR
Claude H. Marks 47

A VALEDICTORY
Kenneth R. Ingham 53

DISCUSSION SUMMARY
Leslie L. Clark 57

LIST OF PARTICIPANTS 65

FOREWORD

It is implicit in the research and development process that unless one has adequate knowledge of the efforts of one's colleagues in the past, one cannot organize one's investigatory equipment optimally in planning for the future. In this respect, persons who work in the domain of sensory research are time-binding, much like their colleagues in literature and history.

This record of a conference assumes the need for documentation in the broad area of research on sensory capabilities and the consequences of sensory impairment. We were pleased to be involved in it for two reasons. The first is that it gave the participants a chance to spell out the requirements, even demands, for documentation in their disparate fields of work. The second is that the papers demonstrate clearly that the problems faced in serving one of the audiences involved are not different in kind from the difficulties faced in serving any other of the audiences discussed in the meetings. These audiences varied from that relatively closed community of workers examining a particular question in research through the entire spectrum to the ultimate consumers of intellective and material products of the research process.

What emerges with great clarity from these discussions is that there is no short-cut to the optimal organization of a data processing system around the needs of a well-defined group of users. One cannot merely adapt another successful solution, for another group of users, and expect great success; an effort must be made to match the capabilities of equipments to the human needs of users in a way which amplifies, and does not merely frustrate, their creativity.

Finally, we are not surprised at the implication in these papers that the need for special interest information resources still exists, and that, in times of crisis, it is to these resources of the nation that we turn for expert counsel.

All these factors, taken together, harden the case for a focussed effort to develop data handling systems adapted to the special needs of the audience to whom their products are ultimately addressed. This is an important inference from this conference; one which we shall keep in mind in our own work. We hope that others will find materials of equal utility in their own work in the papers and commentary which follow.

Leslie L. Clark

PREFACE

This report contains the substance of a conference concerned with documentation research related to human sensory processes held in Pittsburgh on the 13th and 14th of June, 1968. The purpose of the meeting was two-fold. The first was to survey the efforts of information science specialists to date to cope with the problems of exercising some control over specialized literature in several fields. The second purpose was to derive some guidelines to assist designers of information systems. Both purposes were given additional importance by the fact that there has been an increasing tendency for scientists and professionals to utilize information systems for purposes not originally envisaged for them; among these additional purposes are the use of large systems to introduce the unsophisticated worker to the content of a field; and by administrators to follow fields in which they do not maintain an intimate working knowledge.

The conference was co-sponsored by the American Foundation for the Blind, and two faculties of the University of Pittsburgh. The planning of the meeting, its agenda, and the persons invited as participants, were the joint responsibility of the American Foundation for the Blind, the University of Pittsburgh's Department of Special Education, and the Knowledge Availability Systems Center. The administrative details of the meeting, and its on-site conduct, were largely the concern of the Department of Special Education and Rehabilitation.

We are grateful also for the special interest of Edison Montgomery, then Director of the Communications Programs; and Paul H. Masoner, Dean of the School of Education, of the University of Pittsburgh, in the conduct of the meeting. Dr. Ralph Peabody, and Professor G. D. Stevens, contributed invaluable help in planning the conference, and in insuring its smooth running. Dr. Samuel T. Price served as local arrangements chairman.

Midway in the task of transcribing the tapes recorded at the meeting, our faithful and hard-working Sony Model 777/S4 tape recorder failed us. The Sony Corporation in Sun Valley, California, returned it to us in operating condition, and with revisions to bring it up to current specifications for their professional recorders; they also refused to accept payment for this work. We should like to acknowledge our grateful appreciation to the Sony Superscope Corporation, and to Mr. Nathan Tushinsky, its vice president, for this unusual act of generosity. Their kindness and swiftness made our job far easier than it would have been without their help.

Finally, we should like to thank our participants for a particularly lively and spirited series of discussions, and for their cooperation in the preparation of these proceedings. Whatever merits this document may have is a direct result of their concern for its quality and their cooperation in its preparation.

It is hoped that these proceedings may encourage at least one further meeting, one in which the emphasis will be shifted toward the translation into practice of the theoretical concerns of this first meeting. There is much to be done in bringing information systems in human sensory processes to greater usefulness to practitioners. There are also problems of interest to the researcher in information sciences that we can also continue to hope for a maintenance of his interest in these same problems.

A. Kent
G. Stevens
L. L. Clark

EDISON MONTGOMERY: GENERAL CHAIRMAN

There are a number of themes which I think will run through all the discussions. First is the question of how to organize. I've heard of a very interesting decentralized system that is represented by ERIC, a different kind of organization for the selection of the information from organization for distribution. You will hear that ERIC represents the *wholesaler*, with some other organization doing the *retailing*. I think the problem that each of these networks face is size: how broad a scope of material shall a given network deal with? Once there is any restriction in size, then we come to the problem of what should be the relationships among networks. Mr. Marron will mention the "one-stop information center"; this is fine if you happen to be a user with the identification of education, but suppose that you are a physician and there is some discovery in the education field, will you know enough to come to this one-stop information center? We are seeing the development on many campuses, of individual information systems. How do these local systems fit into a network, what are the problems of compatibility, both in the hardware and software systems--these are problems that I think we should all keep in mind as we hear these presentations and I think perhaps we will be guided toward some solution to those problems by what Dr. Baruch has to say.

Jordan Baruch received his doctorate from the Massachusetts Institute of Technology, and his principal career has been with the eminent firm of Bolt, Beranek and Newman. He took an extended leave of absence from that firm to create and operate an "honest-to-goodness" network designed for practicing physicians and hospitals, to serve as president of Medinet, a General Electric Subsidiary. When he left Medinet he returned to Bolt, Beranek and Newman very briefly, and then was attracted by a new organization, The Inter-University Communications Council, EDUCOM, and became president of EDUCOM in February, 1968. It gives me great pleasure to present Dr. Jordan Baruch.

NETWORKS, TECHNOLOGY AND OVERALL FUTURES

Jordan Baruch

In our society one of the most common types of information systems that we have is the discipline-oriented information system, which deals with things such as the information in a particular discipline. A good example of this would be the chemical information system (the chemical literature and abstracting and indexing service). One can view these as part of a model: we have the bibliographic information which is very important; the retrieval information, the locator information, awareness information. It is awareness information put there by the same kinds of people who author the papers: people who do abstracting in chemistry are, by and large, chemists. They abstract them as chemists for chemists. The documents are printed for chemists. The information ranges from the machine readable to the documents. We have the same thing in biology and the same thing in pharmacology.

These are disciplines populated by specialists who are interested in communicating with other specialists in this same discipline, and as a result, their information is difficult to use for the people who are charged with the application of that kind of information. For example, if a researcher were interested in vision he might very well be interested in the role of retinal beta carotene in the eye, in retinal structure, in pharmacology. He would like to be aware of the developments in these fields and a host of others. Medicine is another example. The petroleum engineer is still another with the same kind of problems. He too is concerned about chemical, biological, and currently even pharmacological problems. Clearly, the interest that a petroleum engineer has in the chemical and biological literature is different from the interest that a vision specialist has, that a physiologist has, or that an agriculturist has in the field--and yet they all are caused to tap the same strata of information.

The problem that this poses in a general sense is a national problem. We spend, in this country, a tremendous amount of money on the generation of scientific information on discovery, on research; but we spend very little relatively on the marketing of that information, on its distribution. Our contracts pay page-charges for our journals, we subscribe to several journals, and we now support several organizations like ERIC. ERIC has got something very clever. I could use various terms in phrasing questions; I could say curriculum, administration, learning theory, just as well as other terms oriented to the discipline. ERIC has said, "You know, there are people that have to use these data, in higher education, or adult education; our task is to let one of these organizations, let us say in vision research sink a shaft through the layers of disciplines, and tap off that information which is of interest to it.

Now, how does the vision researcher build up the equivalent of an ERIC center? How does he get into the chemistry-biological literature and combine these things into a working center?

The easiest thing we can do is to look at the specific problems that face us. I am going to assume for the moment that development continues as it is likely to, and that much of the information that we are concerned with in the bibliographic area, the abstract material, is available in machine-readable form. Let's assume that these antecedent groups break down into those that are machine-readable, and those that are visual. What the mission-oriented project would like to do is to establish a paradigm or technique to allow a mission-oriented information center, a procedure whereby it can tap this information, combine it, restructure it in a form that is pertinent and relevant to its own field of endeavor, and distribute it. Let us take a look at the problems that lie in the path of getting from "here" to "there." First we'll take a look at the machine readable text.

In the machine readable case, unfortunately, everybody in each discipline thinks he has the solution to the format of the bibliographic file.

In a brief survey of seven different information files, we found eight different formats, seven of which were known. By "formats" I mean something more significant than the rather trivial question of whether they used descriptor words or key words, or whether the word "title" were typed over the title, and so forth. I mean things like: what is the hierarchy of the organization of the file? How are they handled temporally? How is the up-dating process handled? There are many formats. Unfortunately, just as we sometimes deal with our race prejudice by violence, we frequently vent our information prejudice by doing violence in this respect. There is a tendency to think that, since we have a problem with format, what we will do is to impose a standard. If I say this as though I intend it to be an epithet, it is purely because I intend it that way. One of the characteristics of imposing a "standard" is the basic assumption that whoever imposes the standard knows the ultimate organizational form of information in the field.

This is roughly the same sort of arrogance displayed by the people who say the evolutionary process has stopped in Nature. The evolutionary process has stopped even less so in information. So we don't know the most useful form of chemical information for chemists, or biological information for biologists, and it would be gross of us to assume that we can impose a standard on either, or worse, the same standard on both. A group at the National Academy of Science is working on a technique for labeling files. They finally said, "The problem is really not one of standardization, but one of communication." I don't really care what a format is as long as someone has some nonambiguous way of telling me what his format is at any one instant in time. I don't even mind if he changes it from tape to tape, as long as each tape is clearly labeled. That is a relatively easy thing to do (it seems). A project is underway at the moment investigating what is necessary for a language that will permit the labeling of such tapes unambiguously. We have, in theory at least, given such a language the ability to reduce these tapes to a canonical form, such that they could be merged, decomposed, and put together in any design. It becomes conceivable to merge these tapes. If that be the case, then one would liken this visual research center to a group of people who, out of a deference to the Greeks I will call "epistemologists." These are people who are concerned about vision, concerned about visual research, but are more concerned about the origins and the growth of knowledge in the field of visual research rather than the process of visual research itself. They are the substantive epistemologists. Generally you won't find people who have this characteristic--at least not for some academic generations.

What we will find are librarians and researchers, information specialists, who can deal with each other on a friendly noncombatitive basis and form an epistemological team. This group can communicate with chemists, with biologists, with pharmacologists; and can understand the structure and the concepts--by which this information is indexed, why it is indexed that way, what is its order, and how was that order developed.

The searches will have the following characteristics: they will have high retrieval and low selectivity. There will be a lot of false positives; there will be a lot of false drops. They will get a lot of junk they really don't want, but they will *lose* very few that they *do* want, so they are concerned more about thoroughness than they are about selectivity. Let us assume, then, that they can conduct these searches. For a while at least, there will have to be a staff which deals with the non-machine data, doing the same kind of searches in a much more laborious fashion. You will note that this operation is analogous to what the ERIC does in reading the literature and culling from that literature those things important to its area of special interest, and then preparing that literature in a form usable by the bibliographic mechanism.

Our goal here is from this to assemble a memory disc file, an accessable machine readable file that represents a merger of these. Now, I have made it a random access file for the moment, because one thing which I hope this will develop, as I hope ERIC will develop, is the need to encourage the running of an annotated service. This mission-oriented library is, because of its nature, a two-way communication system rather than a one-way. The real value of this library will come after we have merged files on this disc to allow substantive specialists access to those files to do searches, and after reading the documents to which those searches point, to make notations on those files. This is the equivalent of the marginalia we have in our personal books. The annotation of these files does the following:

(1) It takes a bibliographic entry prepared, e.g., by a chemist, which has key words assigned to it, or descriptors, that come from a chemist's thesaurus, and appends to that words that come from an ophthalmologist's thesaurus.

(2) If we have a text presentation system, it permits actual text commentary, so that any individual bibliographic entry grows both in size and in usefulness through the impact of the substantive specialist.

(3) Presumably, one of the things that this technique also does, is to allow the substantive specialist a very powerful stick: he can delete information and this way we simply get rid of the false positives. Note that what we are doing now is taking information that grew out of the disciplines, moving it into a mission-oriented center, and annotating it so that it becomes essentially a mission-oriented library.

Now we have to take the secondary step, and this is to secure, probably in reprographic or micrographic form, the set of documents to which bibliographic file now points. Presumably this will be done by securing them from the discipline-oriented libraries. This brings up the question of cost.

We have a very peculiar situation in this country. Our discipline-oriented information systems are substantially supported by the government, less heavily by the contributors of the information, and least heavily by the users of the information. An analogy might be that the government paid $4.00 toward your toaster, the manufacturer paid $2.50, and you paid the additional $.79 yourself. It is not quite structured the way the rest of our economy is structured. In fact it is metamorphosing. It is metamorphosing in that the prices of journals are rising, which means the user is paying more of the burden. The collection of page charges is getting more difficult, which is the consequence of two phenomena, namely the reluctance of the contributor to pay for the journal, and the reluctance of the government to support the journal, at the contributor rather than the user end. If this be the case, then these people who represent very heavy users, indeed the most motivated users are the most logical ones to tariff, because they are outside of the discipline. It would be nice to provide methods for them to pay a reasonable amount towards the use of this material. This could be done in two ways: *Chemical Abstracts*, for example, is now charging for its machine-readable abstract information, and charging a reproduction license to provide microfilm. This is a precursor of the future. Eventually what this will grow into is a mission-oriented center paying for much of its flow of information to the discipline-oriented centers, having one branch, among the disciplines because, none of these mission-oriented centers is, in fact, free of being a discipline-oriented center as well. There is a large body of visual literature that isn't going to be reviewed elsewhere that this group is going to have to review, process, document, and publish. In that sense it carries a certain amount of discipline activity with it as well. It will have associated with it substantive epistemologists, substantive specialists, a large retrieval system, and hopefully, a real-time remote access system that allows other users, who are not located in the center, to read and abstract those files in machine readable form for local storage; to annotate those files if they have passed an entrance examination; to modify them; and to add to them--and as a direct result to help pay for them. Hopefully, there will be appended to this a network of terminals, which will permit bibliographic research, initial exploratory research. A request can be sent saying, "Send me the items I've covered."

This transition from discipline-oriented structures to mission-oriented structures is being approached on a number of fronts in petroleum, agriculture, vision, and medicine. One goal of EDUCOM is to try to set up a center which would have as its basis the pertinent technology, as well as the ability to do machine-readable searches in the disciplinary files, so that petroleum engineers could use this technology center to set up a "first test" of the petroleum engineer's information center. We could experiment with the hierarchy of information--the ways of organizing information with its own special retrieval thesauri. Once its needs were known in terms of traffic and technology, it would be possible to build a specialized center.

ERIC: A TOOL FOR RESEARCHERS IN SENSORY PROCESSES

Harvey Marron

The basic theme of this conference is documentation research in sensory processes. Most of you already know that the ERIC program collects, stores, and disseminates educational information. It neither conducts nor supports research--except that which is devoted to making the system itself operate in a more efficient manner. ERIC has something to contribute to this conference in two ways: (1) the programs of ERIC play an important role across the whole spectrum of educational research, including the dissemination of information to researchers working in the area of sensory processes; and (2) the ERIC System itself is an innovation in information dissemination systems which is worth studying. Therefore, I am delighted to have the opportunity to address this conference, and I hope by the time I am finished you will agree that a discussion of ERIC is pertinent to our major topic.

At the expense of being prosaic, I will describe ERIC; what it is; how it works; and how we disseminate information. Then we will look at some of ERIC's accomplishments to date and where I think the program is going from here. I would also like to focus on some of its unique features, especially its decentralized nature, and to discuss some of the strengths and weaknesses of the decentralized information system. In my biased view, ERIC is making a significant contribution to the technology of information handling, especially in view of our current interest in information networks.

ERIC DESCRIPTION

ERIC is an acronym for the Educational Resources Information Center. It was formerly the Educational Research Information Center but the Office of Education realized that there were user groups other than researchers sorely in need of information services. There was a need to disseminate research information, but other information as well (e.g., administrative tests and demonstration programs, ongoing research projects, etc.). In order to make it clear that ERIC was not solely concerned with disseminating research information to researchers, the name was changed to its present form.

Currently, the ERIC program is the only national government information program which, by design allows its subject experts to remain and grow in a chosen professional environment. The information products that result are not generated by documentalists nor information scientists, but by subject specialists working in their specific educational specialties. Their bibliographic or documentation duties are therefore adjuncts to their primary career thrusts. We feel that this arrangement has some distinct benefits which we will review later in detail. There are several unique problems which arise because of this decentralized system and those will also be examined.

The ERIC System is composed of four groups:

1. Central ERIC, the headquarters staff in the Division of Information Technology and Dissemination of the USOE which is responsible for coordinating, developing and managing the system.
2. The network of 19 clearinghouses, each focusing on a specific topic or field. A list of the current clearinghouses appears in Figure 1.
3. An ERIC Facility currently operated under contract by the

North American Rockwell Corporation (NR) which provides centralized document processing activities as well as computer, file management and technical services support. NR prepares the magnetic tape which the Government Printing Office (GPO) uses to print the monthly issues and the cumulative indexes to *Research in Education* (RIE). NR is also responsible for preparing other major ERIC products generated from the ERIC computerized files.

4. The ERIC Document Reproduction Service (EDRS) currently operated under contract by the National Cash Register Company. EDRS sells the full text of the documents cited in RIE or other catalogs in microfiche or hard copy at nominal cost.

ADULT EDUCATION
Syracuse University
Syracuse, New York

COUNSELING AND PERSONNEL SERVICES
University of Michigan
Ann Arbor, Michigan

DISADVANTAGED
Teachers College
Columbia University
New York, New York

EARLY CHILDHOOD EDUCATION
University of Illinois
Urbana, Illinois

EDUCATIONAL ADMINISTRATION
University of Oregon
Eugene, Oregon

EDUCATIONAL FACILITIES
University of Wisconsin
Madison, Wisconsin

EDUCATIONAL MEDIA AND TECHNOLOGY
Stanford University
Stanford, California

EXCEPTIONAL CHILDREN
The Council for Exceptional Children
Washington, D.C.

HIGHER EDUCATION
George Washington University
Washington, D.C.

JUNIOR COLLEGES
University of California at Los Angeles
Los Angeles, California

LIBRARY AND INFORMATION SCIENCES
University of Minnesota
Minneapolis, Minnesota

LINGUISTICS
Center for Applied Linguistics
Washington, D.C.

READING
Indiana University
Bloomington, Indiana

RURAL EDUCATION AND SMALL SCHOOLS
New Mexico State University
Las Cruces, New Mexico

SCIENCE EDUCATION
Ohio State University
Columbus, Ohio

TEACHER EDUCATION
American Association of Colleges for Teacher Education
Washington, D.C.

TEACHING OF ENGLISH
National Council of Teachers of English
Champaign, Illinois

TEACHING OF FOREIGN LANGUAGES
Modern Language Association of America
New York, New York

VOCATIONAL AND TECHNICAL EDUCATION
Ohio State University
Columbus, Ohio

Figure 1. Current ERIC Clearinghouses

ERIC CLEARINGHOUSES

The foundations of the ERIC System are the decentralized centers of subject experts--the clearinghouses. An organization that would like to operate an ERIC clearinghouse has to establish its competence in the subject area in question as a necessary condition for consideration. Selecting and processing documents for announcement in RIE is only part of the responsibility of the clearinghouses. The professional staff of each clearinghouse directs the preparation of selective bibliographies, reviews of research, and interpretative summaries. They make arrangements with those concerned professional organizations to disseminate the material through professional journals, as part of national conferences, and through other communication channels. Major bibliographies and review papers developed by clearinghouses are put back into the system by announcement through RIE and made available through the EDRS.

The clearinghouse products include newsletters, bulletins, research reviews (ranging from relatively short to exhaustive), interpretive summaries, specialized bibliographies, and journal articles and monographs on topics within the area of their subject competence.

ERIC PRODUCTS

The current ERIC products are of three kinds. First, *Research in Education* (RIE), a monthly, is the principal and continuing outlet for the literature of education. All significant documents added to the ERIC collection are announced in this publication. An abstract is provided for each document, along with identifying material. Entries are indexed by author, by institution, and by subject. RIE also announces all new research projects awarded by the Office of Education. They are indexed in a similar manner. The publication can be purchased from the Government Printing Office at $21.00 per year, domestic; $26.25 foreign; and for a single issue, $1.75. The full text of items cited in RIE is available from EDRS in microfiche or hard copy form at attractive prices. The microfiche sells for 25¢ per fiche if all items in RIE are purchased through a standing order. Hardcopy, at about 70 percent of full size, costs four cents per page.

Second are the special document collections in the areas of current high interest and need. Generally, a catalog containing abstracts and indexes provides browsing and search capability. The individual reports, when identified, can be purchased from EDRS in microfiche or hard copy--as with reports cited in RIE. The RIE price structure is in effect for the full text of the reports cited in these catalogs. A list of the special document collections available through ERIC is shown in Figure 2.

Third is the dissemination of selected documents which it feels are most responsive to the needs of the portion of the educational community it serves. These range from newsletters to exhaustive research reviews. It is important to remember that processing documents for announcement in RIE is only part of the function of ERIC clearinghouses. The professional staff of each clearinghouse directs the preparation of bibliographies, research reviews, and interpretive summaries. They make arrangements with concerned professional organizations to disseminate material through other professional communication channels. Major bibliographies, review papers, or other publications developed by clearinghouses are cycled back so that they appear in RIE.

The clearinghouse products include, but are not necessarily confined to newsletters, research reviews, bulletins, analytical reviews, ranging from a short "snappy" item of two or three pages to exhaustive monographs, interpretive summaries, specialized bibliographies, papers at professional meetings or journal articles on topics within the areas of their subject competence.

LEXICOGRAPHY

A Thesaurus of ERIC Descriptors, now being sold by GPO, has been developed and is being continually

PUBLICATION	GPO ORDER NUMBER AND PRICE	PRICES FOR MICROFICHE*
SUBSCRIBE TO THE BASIC ERIC PUBLICATION....		$.25 per individual microfiche STANDING ORDER price for all documents will vary with each issue, but will be calculated at a rate of 8.4¢ per microfiche. (Estimate $90 per month average) $490 for entire collection of 3,796 titles January-June, 1968.
Research in Education. Monthly Abstract Journal reporting newly funded research projects and recently completed research or research related reports of interest to the educational community. Contains abstracts of projects and reports indexed by subject. Author or investigator, and institution. Please send subscription orders separately from orders for other publications listed below.	Yearly Subscription Domestic $21.00 Foreign $26.25 Single Issues $1.75	
FIRST RESEARCH IN EDUCATION SUMMARY, 1967....		
Research in Education, 1967 Annual Index, Reports	(Order by title) $3.25	$.25 per individual microfiche
Research in Education, 1967 Annual Index, Projects	(Order by title) $1.50	$370.00 for entire collection
Cumulates indexes of first 14 issues of RIE, Nov. 1966-Dec. 1967, 2,349 titles		
A VALUABLE HISTORICAL COLLECTION....		
Office of Education Research Reports 1956-1965, Resumes Abstracts of approximately 1214 research reports received by the Bureau of Research before the start of publication of RIE.	OE-12029 $1.75	$.25 per individual microfiche $280.00 per entire collection
Office of Education Research Reports 1956-1965, Indexes Indexes reports by author, institution, subject, and report numbers.	OE-12028 $2.00	
ASSISTANCE FOR THE DISADVANTAGED....		
ERIC Catalog of Selected Documents on the Disadvantaged. Number and Author Index. 1746 Documents dealing with the special educational needs of the disadvantaged.	OE-37001 $ .65	$.25 per individual microfiche $230.00 for entire collection
ERIC Catalog of Selected Documents on the Disadvantaged. Subject Index.	OE-37002 $3.00	

*ALL DOCUMENTS ARE ALSO AVAILABLE IN FACSIMILE HARDCOPY REPRODUCTION AT $.04 PER PAGE.

Figure 2. Special ERIC Collections
(Part 1 of 2 Parts)

Figure 2 (Continued)

PUBLICATION	GPO ORDER NUMBER AND PRICE		PRICES FOR MICROFICHE*
EFFORTS UNDER TITLE III....			
Pacesetters in Innovation, Fiscal Year 1966. Resumes of projects to advance creativity in education approved during fiscal year 1966 indexed by subject, local education agency, and project number. 1075 documents covered.	OE-20103	$2.50	$.25 per individual microfiche $100.00 for entire collection
Pacesetters in Innovation, Fiscal Year 1967. Same coverage as above for 907 projects completed in fiscal 1967.	OE-20103-67	$2.50	$.25 per individual microfiche $120.00 for entire collection
HIGHER EDUCATION COLLECTION....	Order only from NCR/EDRS		
Selected Documents in Higher Education, Number and Subject Index.	ED 012 110	$3.24(HC) $ .50(MF)	$.25 per individual microfiche
Covers 845 documents. Purchase this catalogue only from NCR/EDRS.			$115.00 for entire collection
COLLECTION BY INTERAGENCY COMMITTEE ON MANPOWER RESEARCH....			
Manpower Research, Inventory for Fiscal Years 1966 and 1967.	OE-12036	(IN PRESS)	$.25 per individual microfiche
Report Resumes and Indexes by Subject, Institution, Sponsoring Agency, Investigator, and Report number, with cross reference index covering 393 documents funded by Department of Labor, OEO, HUD, and HEW.			$60.00 for entire collection

*ALL DOCUMENTS ARE ALSO AVAILABLE IN FACSIMILE HARDCOPY REPRODUCTION AT $.04 PER PAGE.

Figure 2. Special ERIC Collections
(Part 2 of 2 Parts)

expanded under the direction of the Central ERIC staff. An overview of the lexicographic work in ERIC is maintained by the Panel of Educational Terminology (PET). Members of PET are drawn from Central ERIC, other government agencies, universities, professional societies, and ERIC clearinghouses. We are attempting to serve the broad range of subject matter subsumed under all of education as well as the specific needs of scholars with narrow specialization with the Thesaurus and its supplements. Recent experiences have given us cause to be optimistic that this approach will work.

Only established thesaurus terms may be used for indexing or for searching the ERIC files. Proposed new terms must be accompanied by a fully executed Descriptor Justification Form (DJF) and each candidate term is reviewed by a lexicographic team composed of subject and thesaurus specialists. The decentralized nature of the system requires a constant attempt to identify terms which overlap the several clearinghouses' parameters. The centralized lexicographic review and control provides the overall system's balance which may not be accomplished easily at the individual clearinghouse level. Two aspects of the decentralized nature of ERIC pose special lexicographic problems. First, the wide scope of the subject matter calls for a plethora of descriptors, while systems or technical considerations require counter-restraints to keep the number from growing too large. Second, the wide range of interest and professional orientation of users forces the indexer into making difficult decisions as how to apportion the indexing terms allotted per document cited. The indexer may, for example, have to choose between index terms broad enough for the general educational community, or very highly specialized ones for the scholars in a particular specialty. ERIC indexers attempt to achieve both but, eventually, operational or technical considerations force compromises.

Early in our experience, we thought that the generality of descriptors would not allow for precise retrieval, however, only a few difficulties have been encountered.

THE USER COMMUNITY

The ERIC system was originally envisioned, and continues to be, an information service available to all segments of the educational field. However, when one counts the total population encompassed by teachers, administrators, planners, supervisors, counselors, and education students in colleges and universities, the figure is staggering. It is not feasible for ERIC alone to serve the total potential user community. Recognizing this, we believe the information needs of education can best be served by a two-pronged attack. A first thrust is to maximize the distribution and the coverage of RIE and the special collection catalogs and at the same time minimize the cost of the reports they cite. This will make the basic document collections widely accessible.

Certainly the development of a comprehensive announcement service and the means to acquire full documents at reasonable cost are prerequisites for the widespread utilization and adoption of new ideas and practices in education. But it is also critical to develop a multilevel set of resources and organizations to perform the more direct information or consultive services for the user community. Elements of this network, aside from ERIC, that may be built into such a system are:

- The 20 Regional Educational Laboratories, which are agents for installing new programs in local school districts.
- The 9 Research and Development Centers which have change-agent roles in various facets of education.
- The 14 Instructional Material Centers which assist those engaged in educational programs for the handicapped.
- The 56 State agencies whose change-inducing responsibilities lie within their respective states.
- The many information, resources and planning, centers funded under various legislative

authorities which are functioning as local "One-Stop" information centers.

In summary, ERIC can serve as a "wholesaler" of information products and services while the groups referred to above can be the "retailers" which can form the essential interface with the users of the products and services.

DECENTRALIZED SYSTEM

ERIC is, at once, a centralized and decentralized system. Unlike most of the large, well developed governmental information programs, ERIC does not gather all of the participating subject specialists into one building or even one contiguous area. Contracts are made with appropriate organizations to operate clearinghouses only after they have clearly demonstrated their subject capabilities, regardless of their geographical location. In effect, subject oriented personnel are encouraged to develop their professional fields. Consequently, the subject competence in ERIC is widely dispersed. Every effort, however, has been made to centralize document processing, computer activities, and management functions in either Central ERIC or the ERIC contractors.

Several advantages of this arrangement are:

(1) the subject oriented staff is encouraged to grow in its familiar professional environment; (2) the specialized document collections at the clearinghouses are not fractionated but built upon and enhanced; (3) the subject experts can be, and are, used for both documentation and analysis functions; (4) the centers of highly specific subject competence are encouraged to develop, and are available for consultation and professional guidance; and (5) the professional staff in each ERIC clearinghouse provides excellent interfaces with specialized professional groups in education, since they are, first, members of their educational communities, and second, documentalists.

The decentralized information system does have disadvantages, some of which are: (1) the lexicographic work (from descriptor generation through usage via indexing) is made more difficult and laborious because of the extended lines of communication; (2) the costs for all operations, because of decentralization, tend to be higher; (3) there is an increased possibility of simultaneous, duplicative processing by two or more clearinghouses; (4) it is difficult to maintain centralized bibliographic control over all of the clearinghouse satellite files (i.e., material not incorporated in RIE but retained by the clearinghouses for local use); (5) the scarcity of information science and documentation experts which makes adequate staffing at each of the clearinghouses especially troublesome; (6) communication at all levels concerning schedules, guides, goals, etc., is difficult and requires a considerable allocation of staff resources.

ACCOMPLISHMENTS

After approximately three years of operation, ERIC has done the following:

1. Developed an announcement mechanism for the report literature of education (RIE) and made provision for ready, inexpensive access to the full text of reports (EDRS).

2. Prepared and disseminated several special document collections in areas of high interest and need.

3. Established 19 clearinghouses, among whose responsibilities are the identifying and processing for announcement the significant documents in their subject specialities and the generation of documents to define user groups in education.

4. Mobilized the information dissemination forces in education to not only disseminate, but to utilize the information products, tools and services currently available.

5. Developed a thesaurus of educational terms which apparently

has been adopted as a world wide standard.

6. Tested and tentatively shown the advantages and the workability of the decentralized information system concept.

7. Developed an announcement mechanism for the periodical literature of education.

FUTURE PLANS

While we will attempt to improve all aspects of the activities we have underway, there are some areas which will get special attention in the next several years. First, we will stress the development of the one-stop information centers. Their roles will include development and maintenance of "front-line" document collections, providing search and reference services at the local level, analyzing, synthesizing, interpreting, and selectively disseminating information, promoting the installation of new practices and providing "feedback" to the initiators of information services and products for their guidance. Second, ERIC will actively pursue the development of an on-line, remote access search system to bring the ultimate user into closer proximity with the document collection. Third, the ERIC clearinghouses are being urged to turn their energies toward information analysis and timely generation of interpretive summaries or digests which can be readily used by the practitioner. Fourth, a strong interest in acquiring the ERIC files on magnetic tape has already developed. Should this demand continue, in spite of the plans to develop an on-line, remote access system -- arrangements will be made for their systematic distribution.

CONCLUSION

Now that we have had a look at the nuts and bolts of the ERIC system, I would like to review our objectives. ERIC attempts to provide all sections of the educational community (practitioners, administrators, researchers, and the partially interested user, such as parents or clergymen) with the information they need to more efficiently and effectively perform their assigned tasks or simply to "keep up" with recent developments. We do not believe that a single monolithic system, no matter how large or well funded, can do the whole job alone. Our approach is to provide librarians or data-bank managers with as many data and information management tools as possible. This obviates their expenditure of resources to acquire these data or the tools for their efficient management and allows them to perform service at the local level. Recognizing the constraints of budget and manpower -- and the enormity of the task -- I think things are going very well.

In my view, meetings like this are very constructive in that they encourage dialogue and cross talk among the information "wholesalers," the librarians, data bank managers, and users. It is inconceivable that truly responsive information programs can be devised without such a continuing exchange of ideas.

RELATIONSHIP OF LARGE SYSTEMS TO SPECIALIZED INFORMATION CENTERS

Joseph Leiter

EDISON MONTGOMERY:

Dr. Leiter is going to talk to us about a different cut of this problem, namely the relationships of a specialized information center to a large information system. To some extent, perhaps, Dr. Leiter is on a very distinguished scale of prototype of the epistemologist that Dr. Baruch mentions. He has had a very distinguished career as a researcher in the field of biochemistry, having received his doctorate in that field, and has done research in carcinogenesis, pharmacology, experimental therapeutics, and cancer chemotherapy, but like what happens to so many eminent researchers, the administrators got hold of him, and he became chief of the Cancer Chemotherapy National Service Center of the National Cancer Institute of the NIH. Having been bitten badly by the epistemology bug in 1965, he then became associate director of the National Library of Medicine. So, may I introduce Dr. Leiter as a distinguished biochemist, a *competent administrator, and an outstanding librarian*, Dr. Leiter.

JOSEPH LEITER:

I'm afraid I have to deny all these characterizations. I'm neither a distinguished researcher, nor a competent administrator, and I'm certainly not a librarian. However, I deliberately left out my capabilities as an information specialist. I do have some papers on this. Some time when I was an administrator in the dim distant past I was asked to take a course for managers in information and computer technology at the Department of Agriculture, and I think my administrative duties forced me to cut at least half of the classes. Nevertheless, I got my diploma, and so it's tucked away in my desk somewhere. I have to bring it forth now and then to demonstrate my competence. Really, the only thing I'm competent in is information science, about which I know very little. However, I would like to keep that at an amateur status, because I really think I ought to be more concerned with the information.

Basically, what has troubled me ever since I became an epistemologist(!) is that most people seemed to be concerned with the framework of the problem rather than the content of the problem, and I was very glad when Dr. Baruch talked mostly about the shell of the system, i.e., the content. I think this is refreshing; I think we would get more useful information.

I have an issue with Professor Kent on this; he is not going to be my P.R. man in this area, because he listed a title for my paper which didn't even have my product listed: MEDLARS didn't appear once in the title. The other issue I have with the title is that it characterized MEDLARS as a system, by implication, and the centers are characterized as specialized information centers. I'd like to make a distinction between these things, and I'd like to refer to MEDLARS as an information center too. I visualize a system as being the shell and the framework, and the center as being concerned with the content of the information, and I like to think that we are concerned with the content of the information. I think there are a lot of systems that could give us the same content; I think we could operate the MEDLARS system on something besides a Honeywell computer, somebody else besides G.E. could have designed it, and it would have worked well if the proper combination of epistemologists existed, as did exist at the Library, to get the thing to work. I think all these are perhaps haphazard associations with the content of the information, and that they are really much less important.

I think, in terms of specialized information centers and large information centers, Dr. Baruch clarified a point which is very important to emphasize. In our thinking about specialized information centers we got things a little bit confused. We think in terms of specialized information centers, like the Vision Information Center, or Metals Information Center, and so forth. Dr. Baruch put his finger on it: the specialized information centers are the discipline activities, and the mission information centers are the ones who are generalists; they have to use all the information from the specialized information centers, so that the Vision Information Center is not a specialized information center: it is a general information center, and uses specialized information from a variety of areas and activities. This is the way I like to think of it. Let me be a little more positive about it, and state that the field of medicine is truly a mission-oriented activity, and it's no different from vision research. The only difference between the MEDLARS system and the other systems is the variety of customers we have. Our customers ask for no more and no less than the specialized information centers; we have biochemists coming to us in an area of cancer, and they don't want a "first cut," they want what they need. We have to be able to satisfy the customers who come to us for bibliographic information, just as though they have come to a specialized information center.

As I'll develop later on in some of our products, you can take as broad approach as you want to the problem of information activities, and you can encompass all the disciplines and all the areas in the MEDLARS system. In Dr. Baruch's terms, I thought I would characterize the mission information centers in terms of "mish-mash." We have to take a lot of discipline areas, mix them together in what looks like a random access disc, but again, this is just a computer's way of organizing sometimes haphazard information into a random universe so you can get at it a little more efficiently. It doesn't necessarily characterize it differently.

This has all been preliminary and introductory; now let me clear this up from the point of view of the paper on ERIC. ERIC is described as a young system. I suppose the MEDLARS, or NLM, system can be described as an old system, or if you want to be generous, as a mature system (if you want to be less generous you can say it's showing signs of arteriosclerosis), but it depends on what sort of service you have received from us which determines whether you say we are mature or we're arteriosclerotic. In any case, mature or otherwise, we, like everybody else, have the same starry look in our eye. Our system is quite old; many think of it as dating back to when MEDLARS was first instituted, in 1960. From a computer point of view it is old. Actually the MEDLARS system is a little over 100 years old. John Shaw Billings, the first director of our library, (which was then the Army Surgeon General's Library) was a war-weary physician in the Civil War. He decided in the late 1860's that he was going to collect and annotate, and prepare bibliographies of all the world's literature in medicine. And he did it for many years. But, like so many other brilliant people, he missed the boat, because in the 1880's he met a young man at a tea in Washington who had the job of assimilating all the data and information of the 1880 Census, and he was having his problems. John Shaw Billings made a sensational suggestion to this young man, saying, "why don't you take pieces of cardboard, and punch holes in them, and use these for sorting out the different kinds of information?" The young man proceeded to do this, and also proceeded to get involved commercially in what is now known as International Business Machines. Whereas, poor John Shaw Billings never thought that in the bibliographic field he could punch holes, too; and it wasn't until 1960 that we first got rid of our barnacles and thought of punching holes in cards for our MEDLARS system.

I thought I would limit my discussion to the bibliographic aspects of the problem. We run a large library. In the field of medicine, as most of you know, the problems of the kinds of documents and the information put in is a little different from what it is in other fields, and we deal very heavily with serial literature; we have aspirations of getting into monographic literature; but we don't have the manpower to do this. But our input, our data store is well defined in terms of information, and so we have different problems than do people in the field of education or technology as far as getting documents to the people, although we've had an active bibliographic program for over 100 years. We began to get it computerized in 1960 and the MEDLARS system (Medical Literature Analysis System) has been operational since 1963, providing a variety of activities. We input about 200,000 articles a year, from about 2,400 serial publications throughout the world's literature, about 45 percent in the English language, and we put this in machine readable form. We use a computer and we use a lot of other things with it, and we come out with products. Over the past four years, our file has been growing steadily. Variation in growth is deliberate; it reflects the exigencies of keyboarding and indexing capabilities; our capabilities to perform fluctuate considerably, but wild fluctuations don't show up because they are quite rapid. At the end of this fiscal year we should have more than 850,000 citations in our file. We have passed the point at which some information specialists said we would fall apart; I think that point was between 500,000 and 600,000 citations.

Although we have an old system, the economics of the time didn't permit us random access, so we have a tape-oriented system with batch processing. We have been able to survive and to maintain services for a substantial period of time, and we provided output more or less on a timely basis; we have crises here and there, which I'll discuss later. Among the principal products that we get from the MEDLAR system is a monthly *Index Medicus*, which is a bibliographic tool for about 16,000 to 18,000 citations a month. We accumulate these in a catalog which is issued biweekly, quarterly, and annually. We accumulate these successively in the machinery to perform a number of other ancillary services: medical reviews, recurring bibliographies, demand searches, and recurring demand searches.

The recurring bibliographies illustrate the wide variety of specialized information services which we provide in a visually readable form, which take slices of information out of the store and make them available to specialized groups. There is a wide variety of specialized interests; e.g., we have one on artificial kidney, surgery of the hand, rheumatology, dental literature, nursing, toxicity, diabetes. These take "thick cuts" and "thin cuts" of the mission-orientations in medicine. A good system ought to be able to do this; the thickness of the cut should be determined by the setting on the microtome, rather than by the sharpness of the blade, if I might use the analogy. Most of the users of these bibliographies are getting the kind of cut which Dr. Baruch has spoken about.

Another set of products is in the development stages and will probably go into publication in the 1968 fiscal year. We go from neurosurgery to cleft palate, to dermatopathology, and even to something as broadly based as nutrition, or cancer chemotherapy. Thus, within the data store of MEDLARS we are performing several specialized information services, for all purposes, and I think this is the kind of "cut" of information through a data base that Dr. Baruch referred to. This is basically a repackaging activity, using the skills and the capabilities of the system to make the packaging as well suited to the group as is possible.

These bibliographies are oriented to broad user groups: anywhere from a few hundred (surgery of the hand), to a few thousand (endocrinology), to as many as 50,000 (dental literature). But that's not enough; that doesn't satisfy all the users. I think it illustrates the fact that there really is no such thing as a narrow mission-oriented group, and I'd like to cite a few examples. As the bibliographies are developed by subject specialists, they naturally consult with the people within their group (and professional societies and groups); they all go through a very similar pattern of information requirements, and I think this is the dilemma of the information specialist, if I may sit back and talk as a user. By the time you get through satisfying every need in neurosurgery, you are back to *Index Medicus*, because you have involved biochemists, physiologists, pharmacologists, and people interested in a large variety of specialities. To satisfy every person on that user base means that we have to provide them with the total base. The same thing is true of anesthesiology. The only reason that you can get a practical bibliography of this kind is that the user group, from this point of view, says that for general dissemination of this type of product, you have to set your targets at a certain level of information, and you assume that for other types of information one uses other bibliographic tools. Even so, they still don't satisfy the individual user. As part of the general plan of bibliographies we also have tailor-made types of uses a demand search or bibliography. Individual investigators can come to us with their research problems and ask us to develop a bibliography for them which will satisfy their own research needs. Since 1964 we have been providing individual searches for people which give them, more or less, degrees of gratification, not necessarily due to the largeness of our system or the smallness of our system, but to our skills in being able to elicit the information from it.

The true test of how good a system is, is whether or not people continue to use it. We are gratified to see we have gone through a steady continuing growth. It isn't the usual graph of probability, in which there is a big rush and drop-off of enthusiasm when people realize that this isn't the milennium. A steady growth has put an increasing demand on us. Both domestically and from two active international centers, the demand runs close to ten or eleven thousand searches a year now. I'll give you some information later on as to the cost, so you will have a better idea of what this involves.

As part of making the service available in 1964, a decision was made to decentralize some of the demand retrieval. We have set up a number of centers. At the present time we have two additional centers, one in the National Lending Library in the United Kingdom, and one at the Karolinska Institut in Stockholm, Sweden. There have been substantial lapses of time between when we said we would establish a center, and when we became operational. I think the interesting thing here is the number of people available to formulate searches; as little as two years ago this was not a large operation, for one or two changes in personnel could limit our capacity seriously. We are happy to say that, as of this year, we have enough people disseminated who are well trained, that we won't go out of business because somebody doesn't like us and gets another job.

I deliberately did not indicate what these people do; some centers just formulate searches and some send them to the National Library of Medicine for processing, and some do their own processing. We do it both ways and it is quite obvious to us that to centralize processing of many of these things is impossible.

Costs are not trivial, although they are not very high. Dr. Frank Rogers, one of our former directors, published figures on what it costs to process a demand search. During the first year of operation the costs were about $120, which was equally divided between machine costs and/or manual labor. There was a substantial drop in labor cost during the second year of operation, and a somewhat smaller drop in the cost of machine operation. This cost, of course, was understandable, particularly because as we began to operate an entire network we began to switch workloads so that when the National Library had available capacity, other centers got searches from us; it is just a matter of the available machine time. There was a progressive decrease in the amount of time spent on each search.

I expect that the labor cost of processing a search will probably level off somewhere between $20 and $25. We have noted that other centers learn a little from the experience of the first, in that they are able to operate to reduce their costs in a shorter period of time. We would expect them to reach a plateau after about two years of operation.

Much of the development over the years didn't come without trauma and troubles. That has to do with the system, rather than the content. Unfortunately, what people think is going to happen doesn't really happen. Sometimes the systems design is off base from actual experience. At least so far as demand searches are concerned, the way G.E. designed the system, the average practitioner will ask a simple question, a few terms from the thesaurus would be put in, and we would be able to load the machine. Instead of having simple questions, they turned out to be very complex. They not only involved a complex manipulation of the vocabulary, but frequently the search analyst had to go do reference work to find out what the subject matter was about before he could formulate a search.

As a net result, instead of being able to formulate eight or ten searches a day, they are struggling to execute three a day, and many start off handling one a day. As to batch size, because it was postulated on the basis of simple questions, it was estimated to be about 75 in a batch, which would mean that with high speed search techniques the search would be efficient, and the cost of the machine time would be very small. As it turned out, when I came to the Library two years ago, the batch-size was only eight, because of the complexity of the searches. By modifying search strategies, we were able to increase batch-size, and we are now running batches of 40.

The system design postulated a static system with a perfect vocabulary, and very few changes anticipated in it. As it turned out, instead of having a nominal 50 changes in the vocabulary of 6000, we turned our vocabulary around 10 percent every year; and we restructured it, so that the job of vocabulary and file maintenance became a formidable one. Recurring bibliographies were supposed to be easy: just construct a strategy with help from others, and then run the request. Instead of turning out 50 simple bibliographies, the best we could manage in four years was ten. We were only able to produce the ten when we changed the system.

As far as the static aspects of the searches are concerned, which was somebody's dream: instead of having an annual review, searches are modified as one gets experience almost every time a bibliography is produced.

Decentralizing seemed to be a fairly simple matter; one writes contracts and gets people to do things for you. As it turned out, between January 1964, and April 1968, we just began to roll. We went through a great many problems; someone decided that a high-level language was important, and would make exchange compatible, and we almost went out of business because the processing time would be impossible. By the time we straightened that out, we decided that centralized processing might not be such a bad idea, because we were forced to do so. We established centers, had search requests come in to them, and we processed them centrally. We protected ourselves from an overload. By mid-1967 we decentralized much of our responsibility for searches, but did more centralized processing, and we began to get teletransmission of the search requests.

This is essentially a painful historical account of how hard it is to start a simple network, using simple tools.

As far as the problem of search strategy development and machine processing is concerned, the tempo increased early in 1966. We made what seemed to be an innocuous change in our vocabulary. We decided that subheadings were a good thing. So we introduced subheadings. No one took the trouble to see what the implications were. As a result, our processing time increased until we were spending 45 minutes for a search. We ran out of time, but we had resiliency in our system, and by a few changes in the logic of our search strategy, and modification of the computer program, in three months we returned to a normal processing procedure, which gave us rapid processing, and an increase in our batch-sizes. From this point on, our processing time follows the length of the file we search. We think we can extrapolate this to our next generation of computers.

Up to now I have considered only search time. There is about 50 percent more time added for report generating, for developing the printout, so each estimate should be increased by 50 percent to get the total time.

Whom do we serve? Since management data is one of the things our systems do not provide us, we do it by the "check-off" system.

We are pretty well spread among a variety of users. Government is a heavy user; half of the requests are from the NIH. Universities with their research orientation represents another group. We have a large number of users from hospitals which are associated with universities, such as Peter Bent Brigham, Massachusetts General, and University of Pennsylvania Hospital. We tallied noninstitutional physicians which, for simplicity, we assume to be general practitioners. We found they were not heavy users of our system. And industry is a very small user of our system. We have a very small "miscellaneous" category, including lawyers, engineers, and sanitary engineers. An analysis was done on a sample size of a little over 900 of our 3389 searches processed during the first five months of 1968.

One of the bones of contention between specialized information centers and large centers like our own is that our vocabulary isn't specific enough; it cannot satisfy the special needs of users. We tried to make a special analysis of one of the specialized information centers in the NINDB network; we worked through the complete thesaurus of over 1700 terms, out of the Parkinsonian Center thesaurus of over 5000 terms. We made a determination of how easily we could map this specialized thesaurus to our vocabulary. Very briefly, 90 percent of the material is directly mappable; either they represent MESH terms (terms already in our thesaurus) or they represent a one-to-one relationship between nonchemical terms; or they represent drugs or chemicals about which we are developing a mapping activity ourselves for our toxicology information program. There are a few terms which are mappable to two or more MESH terms; this indicates some disagreement between MEDLARS and the specialized information center about the specificity of the term. About 6 percent are not mappable at all. By and large, this looks good, so far as taking a specialized information need in a medical field, and seeing how one can use this as an entry vocabulary. Part of our new MEDLARS system design is to use entry vocabularies that map to the appropriate area of identification.

A large number of the nonmappable terms fall into two categories. (1) Either we have a term not in the system, but should have in it, and that *ipso facto* makes it mappable; or (2) there is lack of definition in the term as it is used in the specialized thesaurus, which means that the specialized center has to define the term first before we can map it. This would require: (1) the introduction of some intermediate term; (2) a better, more specific, definition before it can be used; (3) multiple mapping for a term which is definable only in relation to other items in the vocabulary.

I should like to present just one more brief outline. We, in addition to serving our customers, are also interested in knowing what they think of our service. We sent a general questionnaire to most of them. Many people ignored the questionnaire, others told us how wonderful our service is, without telling us why, and some people said our service is terrible, without telling us why. One investigator said that our retrieval was very bad, because more than 50 percent of the citations were "not relevant"; it turned out that 15 citations were retrieved from a bibliography, and only seven were relevant. Some time ago, we made a serious effort to: (1) examine user satisfaction; and (2) find out why a user was not satisfied. The analysis was set up under the direction of Lancaster and has been completed.

A wide variety of user groups was surveyed. They were oriented in different subject areas, and in different organizational areas; we tried to arrange the study so that there were different types of interactions with the system: some writing directly to us, some in which there was an interview with the search analyst, and some in which the librarian was the intermediary in trying to formulate the search request. We processed 300 such requests. As a follow-up, we asked the user to assess the degree of relevance and recall of these retrievals; and we also sent him about 25 actual documents for him to assess, after he read them. All this was followed up, in turn, by an interview, to find out why the documents were, or were not, relevant. In summary, these were the results. First, we seemed to have hit the golden mean; about half the articles we retrieved were considered relevant. We asked the investigators to give us a list of five to ten relevant documents, and we determined how many were in our data base, and we also determined how many of these we retrieved from a request. We also asked them to determine which items were of major importance and which were of minor importance; and we found that we did a little better on items of major importance. This was based on a sample of 6000 articles for precision determination, and almost 2000 articles were evaluated for recall.

We also determined when there was a failure in precision and in relevance and why the error occurred. Vocabulary is not a major cause of recall problems. Twenty-five percent of the recall failures occurred because there was either inadequate communication between the researcher and the search analyst, or because the researcher did not explain his problem well enough. There are similar results for recall.

We would have guessed since we were medically oriented, that we would be strong in the medically oriented areas, and weak in the basic sciences. Surprisingly, we did best in the sciences; we did poorly in disease categories. We expected to do poorly in behavioral sciences, because the concepts are both vague and controversial, and authorities disagree. And we were surprisingly weak in describing such things as how to identify perforation of the gall bladder. These are examples of things you discover doing diagnostics on a vocabulary, instead of using "the numbers game" of "how many do you recall" and "how many do you not recall."

We have also had problems with indexing. We always claim, as does every bibliographic service, that we don't evaluate articles; we simply store what is given in their content. But this is not true; we make judgments by assigning values to the journals, by saying, for example, "Well, this one is a journal that one should index in depth; that one we should try and minimize the number of terms we use." What we are really saying is that one publication is more important than another. There is one other way in which we indicate importance. In *Index Medicus* we say that we cannot afford to print every tag, so we only identify "the most important concepts," and these are the ones on which we put "print" headings. We average 2.6 terms per print heading.

We analyzed these failures in terms of how the system behaves when we let the indexer use whatever terms he needed; when we told him to restrict the number of terms to four or five; and how we did in terms of "the most important concept." Restricting the number of terms does not improve performance. In fact, one does much worse. But if "important concepts" are deliberately selected, one does something by way of improving precision, although you may lose recall, since less important concepts may be of significance to the author.

The depth of indexing is important only for the manner in which it is used, not in terms of the number of terms used; thus 2.6 terms is probably more effective than four or five terms.

The other analysis we did was to show how well the different centers did. There are wide ranges of precision and of recall, depending on the bias. There are some who emphasize high recall who do poorly on precision and follow the general relationships among precision and recall. We have ranked them as to recall and precision. The NLM seems to be in the middle of the ranges, again, and represents a median approach. There is wide variation which suggests that, based on strategies employed, one can raise the rank in any desired direction -- to improve either precision or recall, depending on what the needs of the individual may be.

We did another experiment, independently of this last one which I have reported, concerned with search strategy. It is said there is only one way to formulate a search -- my way. So we decided, both in our regular workshops and in our own searches, to take a few live, controversial searches, and ask many people to formulate search strategies for them. We derived a "spatter pattern." We can get precision and recall retrieval which can vary between zero and one hundred percent, in one type of search. In most searches, the variation is close to the 50 percent range. There are wide variations in retrieval, depending on the search analyst. So this system is, like any intellectual system which is highly dependent on content, on how the strategy is developed, on vocabulary limitations and computer design. It is not just that a small vocabulary can be good, and a big one better, but rather that one can do poorly or well, depending on how well one develops the strategy with which one searches with it.

I have tried to review the problems and processes of a mission-oriented information center regardless of size. With properly trained systems people who can effectively negotiate a transaction with a customer we can ultimately improve services.

SWITCHING AND TRANSFER PROBLEMS

George Howick: Director of the Office of Technology Utilization, NASA

Mr. George Howick, Director of the Office of Technology Utilization, NASA, had an automobile accident on his way to the conference and was not able to present his paper.

EDISON MONTGOMERY:

Mr. Howick was going to speak on the Technology Utilization Information System of NASA. You may know that NASA spans almost all of the sciences and engineering; the NASA contractors and subcontractors, and the work done by NASA scientists and engineers themselves, directed toward the space effort, seems to offer opportunities for application of their discoveries to what the NASA people call "non-aerospace industry." They were troubled with the problem of how to get discoveries in the hands of industry to benefit the general economy.

I do think that it is worthwhile to hear a little about this system, and we are fortunate that Professor Allen Kent is also the Director of the regional dissemination center of NASA located here in Pittsburgh. I have asked him to speak briefly on the NASA system.

ALLEN KENT:

Having invited George Howick to deliver this paper, I think I know what he might have said.

He would have started with a discussion of the data base of NASA, a large file that is being augmented by 4,000 to 6,000 documents a month. This material is indexed in some depth. There is an index recorded on magnetic tape. Its primary purpose is to serve aerospace research and development people, who are working in the aerospace field. A change of horses in mid-stream occurred when regional dissemination centers were set up to serve industry: to serve a different purpose than originally contemplated. This led to the very significant question of how one switches and transfers information generated for one purpose to another purpose, and makes it useful.

This would all have led up to two questions that George Howick would have posed to this conference, both related to one general question, namely: "How does one evaluate the systems and networks?" The first question has to do with the effectiveness of transfer of information from the head of the person, where it belongs; and the second: "How efficient are these systems and networks in providing useful information economically and swiftly?" To restate: First, who really uses the information provided by the systems; how do they use it; and how useful is the product from the system? Second, are these the best systems to provide information to those who are using the information rather than providing information to people who collect information only? We have noted in operating our regional dissemination center that many people who pay for service are in the business of collecting things for their libraries. They like to have things on their shelves. Very often, we find companies stopping the service for a year because they say, "Well, we have enough to read for a year, and we'll start again later." Perhaps this is enough to say to introduce the major question of how does one evaluate the systems, tracing what happens to the information provided.

A SURVEY OF INFORMATION SYSTEMS ORIENTED TO BLINDNESS

Samuel Price

The speakers this morning discussed information systems networks and the technology associated with them; this afternoon, we hope to review the current status of information retrieval in human sensory research.

My purpose is to describe the present state of information services which exist in the United States to serve the information needs of persons who are working, in one capacity or another, in the subject matter area of blindness. To be useful, I had hoped that we would be able to orient the survey toward the idiosyncratic needs of the actual users of the special libraries and information centers on blindness in the United States. This is easier said than done. To determine the needs of users is like the case of the old Yankee lady who found an old recipe for rabbit stew; it started out, on the first line, with the statement, "First, you catch a rabbit. . ." and, over the years, someone had modified this first line to, "First, you find out if what you have is really a rabbit. . . ."

The object of this story is to indicate that if one has never had experience with rabbits, one would never know that once you take them out of their natural habitat, and put them in a stew, it is hard to tell the difference between a rabbit and a woodchuck. When users are taken out of their natural habitat, and only consider the questions they put to you, it is hard to tell their point of view. The needs of this group, we decided to determine indirectly, by asking the operators of the systems (perhaps dangerously) what sort of questions their patrons asked when they requested information services. So, we used a questionnaire as our methodology, and we hope to derive a qualitative picture of the quality, scope, and speed of information services which this hypothetical group of users has.

Having these performance requirements, through the questionnaire, it is possible for us to discern some useful criteria for the evaluation of the systems themselves. We chose blindness as the area that we would pursue, rather than sensory processes in general, for two reasons. One was that we happened to believe that the volume of literature in the field of blindness was of more manageable proportions than it would have been if we had chosen the larger field and survey all of the systems. We also hoped that the information services for this area and group of users are no better and no worse, for any other area of sensory processes, such as deafness.

One more limitation of the survey was that we excluded, or tried to exclude, instructional materials for the blind, or reading materials for the blind, because this eliminates a huge literature that we did not want to be concerned with at this time. We developed a questionnaire with these purposes. We identified sixteen special libraries and information centers in the United States, through directories of special libraries. If the library was in the directory, we included it in the list for the questionnaire; if it was not in the directory, we did not include it in the list. We omitted all those which pertained only to medical collections.

It might be interesting to note that to be listed in one of these special directories, it is necessary for a library or information center to fill out a form, which lists their holdings and the types of services they provide. They return these forms to the publisher of the directory, and this information becomes available to anyone who uses the directory.

The specific purposes of our questionnaire were: (1) to determine the information problems which were of importance to the patrons of these special libraries; (2) to develop criteria for qualitative evaluation of the responding systems; (3) to determine the methods that are currently being used by the system operators to organize their collections; (4) to determine the subject matter expertise, or point of view, of the users of the systems, with the idea that this also would be useful information if we wished to evaluate what was actually being done in the system.

We sent out 16 questionnaires. We got six back. This is not as bad as it sounds, because in each of the questionnaires which did come back, pertinent information regarding each of the purposes was included in full. The agencies who responded to the questionnaire, by type of respondent, were: two professional organizations; one university-affiliated library; two state school-affiliated libraries; and one private school library.

The data from respondents reinforce the analysis given us by Drs. Baruch and Leiter about the problem of compatability of information in the several centers. One of the factors mentioned by Baruch is the matter of bibliographic self-sufficiency; the idea is that a library which is in fact self-sufficient is one which has, or has access to, all the documents which contain all the information which would conceivably be needed by anyone having need of the information. No one has this, of course, at least not internally. If we could increase the compatability between the systems, then it might be possible to improve the external bibliographical self-sufficiency; that is, one would have access to the information now residing in the files of other information systems. Inter-library services, in this kind of situation, are likely to be rather limited. It is not likely that these systems are in correspondence with other systems, and have union lists of holdings, and so on, to share library resources.

Next, we asked respondents to suggest criteria to evaluate the strengths and weaknesses of its non-medical information services on blindness. We asked them to do this themselves, and to make up their own criteria. Then, after doing so, we asked them to rate their own services, in terms of the criteria they had suggested. They did so. Two general categories emerged.

The categories relate to "utilization," that is, the quality of service provided to patrons; and "coverage" of the material in the files. Note that when the libraries were arranged according to these dichotomies, respondents suggest subject coverage criteria more often than they do the utilization of the holdings they have; as a matter of fact, the frequencies are in the relation of two to one.

The respondents considered their own systems weaker in utilization than in subject matter coverage. It appeared that most agencies regarded their services to their patrons less substantial than the subject matter they had.

Another interesting comment is that no respondent suggested any kind of cost/effectiveness criterion as a valid measure of their systems. This may be a crucial criterion.

We next asked respondents to suggest other libraries in the United States which were "very good," in terms of the criteria which they had themselves suggested. They were unable, or unwilling, or did not suggest other libraries that were more substantial in terms of these self-stated criteria.

The next item on the questionnaire was a series of questions about what patrons wanted to know when they sent in a question. The questions most frequently asked pertained to the rehabilitation of the blind or to the education of the blind. It is notable that we are involved in a conference on documentation and information retrieval in human sensory processes; we are funded by a research oriented organization; and we find that the questions we are being asked are nonresearch-related questions.

Below are items and categories which appeared on the questionnaire:

1. music education for partially-seeing;
2. employment of blind persons as teachers in public schools;
3. colleges offering training programs for teachers of the visually handicapped;
4. address list of residential and day-care programs for the visually handicapped;
5. effects of mobility training on the physical education and performance of blind children;
6. information on script-writing for the blind;
7. teaching written expression to the visually handicapped;
8. list of talking books for blind primary age children;
9. demographic information about blindness;
10. state legislation pertaining to the visually handicapped;
11. list of agencies serving the visually handicapped in the U.S.;
12. teacher certification requirements, by state, for teachers of the visually handicapped;
13. states which permit blind persons to be certificated as teachers;
14. where to obtain services for parents of blind multiple handicapped children;
15. research reports on teaching spelling to the visually handicapped;
16. should Latin be taught to blind students?;
17. bibliography on the education of the blind;
18. historical list of all past trustees of the Perkins School;
19. computer programming for the blind;
20. art techniques for the blind;
21. education of the retarded blind child;
22. mobility training;
23. teaching social skills to the blind;
24. studies on attitudes toward the blind;
25. emotional factors affecting the learning of abstract mathematics; (to adolescents blind as a result of RLF);
26. specific documents on research needs in braille.

I took these questions, and broke them up into categories, according to type of question, whether it was a general or specific question, and the frequency of occurrence and the percent. Among some implications we can infer are that raw research findings are not required by the patrons of the information centers that we surveyed; in fact, only 15 percent requested this kind of information. Almost a fourth of the inquiries requested specific administrative data. Questions were predominantly specific; this tells us that information about educational developmental activities is of crucial importance to patrons of these systems; that implementation of research findings is one of the critical needs that the users have told us through the questions they ask.

The subject matter expertise of the users of the systems was easier to provide by the respondents than some of the other information we requested. The professional orientation of those who asked these questions is predominantly education. This somewhat contradicts some guesses we might have made prior to the development of the study. With a heavy emphasis on research, it seemed reasonable that the prevalence of questions dealing with research would have been greater than it turned out to be. The majority of patrons of current systems are practicing educators.

In recapitulation, we studied: (1) the problems that users of the systems had; (2) criteria for system evaluation; (3) methods used by system operators for organizing their collections; (4) the subject matter point of view and expertise of the users; (5) some cost data. The shortcomings of the study make generalizations risky, particularly the small sample size. But some inferences seem to emerge. With respect to the first objective, related to information problems of the users, it is quite plain what the users want the systems to do for them. They want specific items of information in a reasonable length of time, in a form that is useful to them. From the point of view of subject matter, the trend is toward information regarding utilization of research findings in an educational or rehabilitation context. Many of the questions being asked by users are not being answered. Patrons of the system had to go away unsatisfied. Operators of the system are now able to refer clients to other sources of information. Perhaps we can infer some of the reasons for this if we consider the question about evaluative criteria. Libraries exist, of course, to serve their patrons. Yet, traditionally, when library systems are planned there has never been much attempt to build evaluative mechanisms into them. Two facts which emerged demonstrate this tendency; the respondents listed twice as many evaluative criteria in the "coverage" category as they did in the client service area; and the respondents consistently rated themselves less well on service factors than they did on subject coverage factors. Information services cannot be provided to clients with subject coverage. Yet the service factors are also critical, since information service is the one characteristic by which libraries or information centers justify their existence. We talked at the outset about the way in which the libraries were entered into the directory of special information resources and the participants in our study were chosen because their names appeared in that directory. There is an incongruity between what appears in a directory and the responses we obtained with our questionnaire.

In my opinion, the most glaring deficiency was the lack of cost/effectiveness criteria for evaluating the services: how much it costs to operate the library, the price of service to clients.

Although the respondents showed no hesitancy in rating their own services, they seemed reluctant to suggest other information services which they considered to be very strong. Operators of the systems may be unaware of the information resources which are available to serve their clients. Moreover, there may be incompatibility between the various indexing systems used by the respondents to catalogue their collections. Such compatibility would be very useful to operators of these systems. It might then be possible to refer clients with unmet information needs to other information centers; they might also then be able to add inter-library loan service to their repertoire of service strategies.

Sharing library resources depends, at least in part, on compatibility of cataloguing or indexing systems used by the institutions or organizations which wish to enter into this kind of cooperative enterprise. If the compatibility exists, then the union list of holdings becomes a possibility. The findings of the survey would seem to indicate that the compatibility which is needed doesn't exist.

This study is an exploratory probe to suggest trends, to show what exists in present systems, and to indicate the desire for improvement in their capability to handle the literature on blindness. More important, it clearly reveals a need for definitive study.

THE AFB RESEARCH INDEX

James W. Linsner

In work for the blind, as in any other field of endeavor, there is a constant need for information. What others in their field are thinking about and experimenting with is of great interest to engineers, teachers, psychologists, and other professionals. They want to apply the understanding and insights gained through the experience of others to their own work. If they are researchers, it is crucial that they know the history of similar research and keep track of current projects.

How are these people to find the information they need? How are they to learn of the existence of an article or report which may be important to their work? How can busy professionals find time to wade through the thousands of articles, dissertations, patents, government reports, etc., that now appear every month? The answer is, of course, that they can't. The "information explosion" has made it impossible for a man to do both his work and his reading. He usually finds that specialized searching and selecting services can be of great assistance to him.

The International Research Information Service (IRIS) is in a unique position to provide such a service for those interested in research for the blind. The IRIS staff is conversant in both technological and psychosocial research, and can selectively peruse many scientific journals, abstract lists and other acquisition lists each month. Furthermore, the staff is in touch with researchers throughout the world who frequently send in project reports and unpublished memoranda regarding their own work, and who often alert IRIS to the existence of other documents of interest, including foreign language items.

The IRIS Document Center has been collecting and codifying this information since 1962, and is now offering an index to these documents. The new AFB Research Index is in loose-leaf form. An updating service will be provided in the form of quarterly supplements, to be issued on a subscription basis beginning in March, 1968.

As with any documentation system, the basic question is one of purpose. Purpose, more than anything else, determines what kind of information is to be collected, and to what kinds of users it will be furnished and in what form. Our purpose is to supply needed information to researchers, whether they be sociologists, neurophysiologists, or electrical engineers, and to professionals directly engaged in services to blind persons--teachers, consulting psychologists, etc. Thus the documents must necessarily derive from a broad range of the natural and behavioral sciences and technological specialties. What kinds of user need will such a collection satisfy?

We can, I think, distinguish three basic kinds of information needed. The first is usually called "current awareness." This is the need most keenly felt by researchers. A researcher needs to know what is being done by others working in the same field. This information enables him to avoid costly duplication of effort and often provides stimulating questions and ideas from which he can profit in his own research. The second type of need, while somewhat less common, is equally important. We may call it a generic or "state-of-the-art" survey. Both researchers and professional workers occasionally want to go through all of the research findings, both past and current, on a particular subject. The third and least common need is for a highly specific piece of information--a seldom-used technique or formula, etc.

The AFB Research Index is designed to satisfy the first two of these needs. "Current awareness" is, of course, a relative term, and to some, quarterly supplements will not seem frequent enough. For the present, however, the influx of documents is not large enough to warrant more frequent distribution. "Generic surveys" will be particularly fruitful, given the general qualification that the research must have something to do with the problems faced by blind persons. For example, if a user looks under READING, he will not find all possible information about reading. He will find investigations of the reading problems and habits of blind persons.

There is one further type of information that we have attempted to capture in this collection. It concerns certain areas of basic research, such as sensory physiology, sensory deprivation, neural coding studies, etc. Some of the investigations in these fields may ultimately be of great value to those engaged in more directly pertinent research projects, and we are selectively screening this "fall-out" from basic research. A user who looks under SIGNAL DETECTION, for example, will find only a very small fraction of the literature, but he will find, among other things, some interesting reports of comparisons between auditory and tactile or auditory and visual function. This, then, constitutes a second and more limited type of generic survey which can be performed with the Index.

However broad the range of subject matter in the IRIS collection, the anticipated size of the collection is fairly small. Consequently it was decided that for the present and foreseeable future, an efficient manual system for retrieval of the information would be preferable to a more costly mechanized system. We chose a book-form system in which no cards, hole-punchers, illuminated screens, or other pieces of equipment are necessary.

The Research Index is a three-volume loose-leaf set composed of a *Dictionary of Descriptors*, a *Scan-Column Index*, and a *Bibliography*. When a document is acquired it is assigned a serial number and is then indexed using terms from the DICTIONARY OF DESCRIPTORS. The indexing is "natural" in that the terms are taken from the documents themselves. For the most part, they are words and phrases in common use, such as "adjustment," "braille," "communication," "compressed speech," "programmed instruction," "reading machines," and so on. Following each descriptor is a definition and a list of related terms. There are also many cross-references in the dictionary.

Each of the 130-odd descriptors has a corresponding letter code, and after a document has been indexed, these letter codes are recorded in the *Scan-Column Index*. This is the book-form indexing and retrieval system designed by Dr. John O'Connor of the Institute for Scientific Information. It is an efficient coding scheme, and, fortunately, a fairly easy one to use. As an example, to find all the citations pertaining to intelligence tests for the blind, the user merely runs his eye down the 13th column of each Scan-Column page looking for the symbol "t." Every time a "t" is found, he writes down the serial number of that document. When he has finished this process, he has a complete list of references by serial number. If he is interested only in tactual, nonverbal intelligence tests, he must scan for three symbols at the same time. This is much easier than it sounds. A three-term search, for example, can be performed at the rate of about 500 documents per minute.

Armed with his list of serial numbers, the user now turns to the third volume, the BIBLIOGRAPHY, where citations are listed serially. Each citation includes all authors' names, the title of the work, the project or grant number, publishing and source information, and a list of the codes for all terms by which the document was indexed.

While we believe that the Index is quite easy to use, anyone who has difficulty with any part of the system may, of course, request our assistance. If, for example, he is unable to translate his interest into the terms of the Index, he should write us a short paragraph describing his interest, and we will formulate his inquiry in descriptor language. On request, we will assemble short subject or author bibliographies. We will also help users in obtaining out-of-print items and unpublished reports which cannot be found elsewhere. However, for the present, we will not be able to supply the documents themselves.

For the next two years, we will be inviting the criticism of both information specialists and actual users concerning the effectiveness of the Index, and we intend to remain flexible enough to incorporate such changes and supplementary services as are appropriate and are within our physical capability. In this way, the International Research Information Service hopes to fulfill, in one more concrete way, its sole function, which is to provide information to researchers and others engaged in work for the blind.

THE VISION INFORMATION CENTER

Robert Reinecke

I would like to clarify one or two points as to what the Vision Information Center is, and particularly its relationship to the National Library of Medicine. Since Dr. Leiter gave his paper this morning, you may wonder about some of these various things that are going on. I'd like to sketch a little bit how we fit into this picture; or at least the way we imagine we fit into it.

A few years ago the Weinberg Report,* pointed to a need for information centers in the United States. Eventually, this came to fruition, and in fact it was already going to some extent, in the name of the National Library of Medicine. But it was deemed necessary that there should be an extension toward the concept of an information center. This became particularly evident to the NINDB of the NIH. Under their auspices, contracts were granted to several schools to establish some specialized information centers. I'd like to spend the remainder of this time telling you not so much what we are dreaming about doing, but what we are actually doing at the Vision Information Center.

Our relationship to the National Library of Medicine, is a friendly, informal relationship, which is essential between information centers and which lends strength to the entire program. What is this relationship? We cooperatively do the indexing of the articles related to vision. In addition, they're training some of our indexers so they will do a thorough job.

While this (the NINDB information system) is termed a network, it really isn't in the true sense of the word. There are several reasons for this, but technological limits is the most cogent reason. Actually, we had no national hook-up or some giant computer to use, to make this type of system work. The entire system is dedicated to the idea of having an information retrieval system which is extremely fast, and can operate with quick access. The four centers have gone their separate ways. We do meet and talk to each other and try to accept ideas that are best for each center, but it will be a few years before we can really tie them into a cohesive unit, culminating in a more direct relationship with the National Library of Medicine. These centers are different from the centers described by Dr. Leiter, where the MEDLARS searches are formulated. Although MEDLARS searches are formulated at Harvard, it is quite distinct from the Vision Information Center, although we're housed at the same place.

So much for the overall approach. But what are we really doing? How do we justify our business? What are our plans? How do we get going?

Most of you are probably aware that the use of libraries has decreased over the last few years, when you look at hours spent in the library versus the number of students. This was studied intensively by the INTREX group at MIT. It seemes to me that there ought to be a wedding of education, information retrieval, and bibliographic systems in libraries. Most of the things I will say will be designed to support that thesis. While we are working under this thesis, I am not sure that the other three centers believe this is a necessary goal to work toward.

---

*"Science, Government, and Information," The White House, January 10, 1963.

I might give you a little background of my own interest. I've been interested in programmed instruction, and other techniques of learning, and it became apparent that computer system instruction, while not a panacea, offered something a little different. We might now be able to establish a random access means of getting specific questions answered, and also be able to go into a data base at any time. Accordingly, we tried to merge a data bank, namely a bibliographic system, with a computer-assisted instructional format which is random access. This is a rather prodigious undertaking, and one that really won't come to complete fruition for a while, but we believe we're making some significant inroads into the problem.

Let me go back a bit, to when we decided to get into business. We asked the question: What services can we provide the user?

Maybe we ought to do a survey and see what users want from an information center as compared to the National Library of Medicine. One of the things we decided, very early, was to set up a scientific advisory committee. The committee is made up of specialists in ophthalmology, optometry, and education, and we meet at regular intervals. This committee has been very active in helping us to formulate policies. In general, the committee has been favorably inclined toward the goal that I outlined to you earlier.

One of the first things we had to do was to get a data base of some sort. This meant beginning right away to index material. We had no indexers immediately available, and the National Library of Medicine takes three or four months to train any indexers. We contracted with *Excerpta Medica*, and they immediately started providing us with indexed articles. From this we devised a thesaurus which is structured in a rather detailed fashion. The user will have to use this thesaurus in formulating searches. I'll describe it in a little more detail in a moment.

Our input is essentially the bibliographic input to our system. From this, and concomitant with it, we developed a thesaurus which is a hierarchical arrangement, based on numbers. We can redesign it very easily. It is reasonably compatible with the National Library of Medicine and MEDLARS system.

We developed a thesaurus, continually expanding, we use to index our instructional material. We feel it is easily adapted to computer usage. We have modified our computer capability and we're coming up with a system which will be operational in the fall of 1968.

When we looked around the country at the various types of information centers, it became apparent that practically all of them had this "buffer zone" that has been discussed above: the search analyst, or literature specialist, or whatever you wish to call them. Essentially you have a data bank and the user is remote from it. He really cannot get to the data unless he goes through one of the information specialists. Unless you're one of the "in" group, you really can't get to the material and use it yourself; you always have to rely on someone else. It became apparent to us that this was a serious deficiency, as has been pointed out by other participants, and it was also pointed out that the only way you can ever have meaningful use of a data bank is to let the user go right into it, sample it, and see whether or not he likes it. If he doesn't like it he can go somewhere else; but he's got to be able to get into it. In that way, we hope, you'll have a viable product that will go on almost indefinitely. If you really rely on specialists, you're taking a very big chance, particularly with smaller, more specialized units.

Essentially, then, we have a permuted alphabetic sort, where we take everything and line it up in the middle. The most important consideration is really that any synonym has the same index number. The alphabetic numbers are generally compatible with those of the National Library of Medicine mesh system. However, they are not exact--nor do we intend them to be. It's obvious that if we're really concentrating on vision, and all the aspects of it, so eye disease for example is C13, we really have to give this a major heading; whereas in the National Library of Medicine this falls in sensory disease. It also became apparent that we have to be able to index everything. This has been pointed out by others in this conference. If you're going to handle a special eye segment, it is necessary to index everything in the article. Also this is essentially an alphabetic presentation, and the programs are such that we can print this off new whenever we need it -- every three weeks or so, or every week if necessary. The other thing that we want to do is to have it set up on an hierarchical basis, so that that person can be more general or more specific, depending upon his needs at the moment. If he's doing a review of some broad subject, obviously he wants to get everything he can; on the other hand, if he's trying to find a selected bit of information he wants to narrow it down.

After the person locates the numbers he wishes, relating to the term, they simply look at an hierarchical arrangement, and then they can be as specific as they wish. They can then decide whether or not they want to be more specific or less specific. If you were to look through our thesaurus, it would become apparent to you that large segments are omitted. We only include things in our thesaurus that we have in our system. We don't put anything into the thesaurus system if we really have nothing on it in the data bank.

Early in the history of this system we thought we'd better make a prototype to see whether or not this would really work, if people would like to play with this kind of thing, so we wrote a prototype computer system instructional language with the 1401, and played with this. It was not only enjoyable but people became enthused. There's no better way to get people involved than to get a model to work and have them sit down and let them do it themselves.

We also put in some selective bibliography, and if the user wishes, he can have bibliographies which we have readily available. This demonstration program is fairly small (500 frames).

He can also be taught to use the thesaurus. Indeed, we found this part to be particularly applicable now, since we're really getting into making the user go to the thesaurus, and the system quizzes him right off the bat, as to whether or not he has one available. If he doesn't have one, he's out of the running. He's told where he can write for one, or get one, and that's the end of his searching technique. But essentially, we show him how the different parts of the system fit together and how he can make his search more general or more specific. As you come down, you become more specific, and we simply use different hierarchical numbers. This is a very convenient arrangement for synonyms.

The system is simple and straightforward. The user sits down at a console, and indicates whether or not he knows how to run the system, and whether or not a thesaurus is available to him. If he has a thesaurus he's in business, and we can teach him the rest of it. He sits down at a console, and a conversational mode is established with him. He then picks a number, for example D1; and he can add D2, D3, as many as he wishes. After each one of these he's told how many references are available (e.g., 150) and he can ask again and the number comes down. If he asks again, the number may come down to 5 or so. So he can pick any of these that he's satisfied with. If he picks less than 20, we'll type them out for him. If he picks more than that, the first eleven are typed to get him going. The rest will be mailed to him; or he can pick them up at the computing center. If he is a person outside the field of vision research, such as a biochemist interested in enzymes, and he wishes to know whether or not it was in the eye, he might come to it and look it over and quickly realize that he needed something much more specific than "the eye." He got back, let's say, 1000 references, and obviously he needed to go to something much more specific in the eye. He might decide that he was interested in the cornea. But he really doesn't know what the cornea is made of. He can then go to the CAI system and really ask for some instruction as to what the layers of the eye are. This type of system is based on the idea of the user coming in and talking to a specialist to get an idea what you're searching for. This is the idea of trying to provide some information easily available to him, to teach him just the rudiments in the field so he can then make an intelligent selection of the terms available. Then he can quickly switch out of CAI mode and back into a search mode. We do have more elaborate search modes for the experts. You can do "and" searches, and you can do "what not" searches. I'm not interested in them personally, because when I'm using the system we have available, which is somewhat more primitive than what I've described here, I find that "and" searches are really very useful and quite adequate, particularly when dealing with very specific terms. Our indexers tend to choose specific terms which are really pertinent to the article. By October 1968 we hope that this system will really be operational and will be demonstrated in Chicago.

What do we have as far as the staff is concerned? How are we operating? What's the volume of things that we're going to be indexing; what is the scope? Our scope consists of all things medical related to vision, and optics as it pertains to refraction of the eye itself. We are not presently involved with things that were listed in the IRIS project, and we are not heavily involved in the psychology of vision, although we probably will be in the not too distant future. Our scope may be changed at any time, depending on the reaction of other centers. As an example of that, if one of the other four centers covers a particular subject we refer requests for information to the appropriate center.

We are attempting to cover the world literature, which means scanning more than 2500 articles that come into Cuntway Library. We estimate that this will total 10,000 articles per year to be covered. We ultimately will try to cover all books, doctoral theses, and monographs. We indicate on the microfiche what we have available on every indexed document, whether there are any abstracts, where they are available, and letters to the editor. And we have copies for all these things, photographs assembled together, by using a microfiche jacket system.

We only go back to January of 1967 and we'll only add to our system, and make bibliographies available, but we will not go back in time and enlarge our data base in a major fashion.

The general chairman for the second day was Mr. Kenneth Ingham, research associate, Massachusetts Institute of Technology. His background is in electrical engineering and theoretical physics, currently engaged in the development of reading machine system for the blind.

CENTRALIZATION, DECENTRALIZATION, AND SPECIALIZATION

Allen Kent

I thought it would be helpful to step back, and examine one of the issues that's been an undercurrent throughout the first day's discussions: to try to bring the problem to light; to seek some causes for this undercurrent problem, even if they may not be amenable to attack in any quantitative sort of way. Nothing that I say will be wholly true for any particular information retrieval system, but every part of it, I think, will be true for one or another situation. So this won't be quite a tale out of whole cloth, it'll be a tale out of parts of cloth.

First I'd like to start with definitions of a few of the terms that I'll be using, just to be certain that you'll understand my thinking even if you don't agree with it. First, the unit operations that I consider to be a good way of organizing one's thinking with regard to the information retrieval field. I hope later on you'll see how these definitions relate to the theme of this talk on centralization, decentralization, and specialization. In retrieval system, there is a set of input operations that are generally performed, and a set of output operations.

The input operations: The first I'd like to label "acquisitions"; one is not in business in an information retrieval system until one acquires the material that one has to deal with in this field. Second, an "analysis" operation that implies various things for various people; this might be abstracting, indexing, classifying, or extracting--anything that involves the perusal of the material that has been acquired, and doing something with it to identify those things that a retrieval system may operate on, and use, in connection with the services that will be provided later on. Third, because the analysis activity is generally inconsistent with regard to the words that are used by people who perform this analysis, there is another unit operation, "vocabulary control." In traditional librarianship this may involve the use of standard subject headings or catalog headings; in this field it may involve the development of a thesaurus or other type of cross-reference system that is considered to be appropriate.

Fourth, after having analyzed and controlled vocabulary in some way, one has to "record the results of analysis on some searchable medium." This might be on a catalog card in a library, or it might be on a magnetic tape in a computer.

Fifth, something must be done about the material acquired; it must be "stored" in some way. These, then, are the five "input operations," that must be considered before one is ready to contemplate what to do about output.

Let us consider the output functions, of which there are three: First, "questions" are asked to initiate the output functions; question analysis and development of some "strategy for searching" (be this a strategy that a reference librarian uses in searching a catalog, a strategy using an alphabetic index, or a strategy involving the logical capabilities of the computer). Second, there is an operation that involves physically "conducting the search," regardless of what the search mechanism happens to be. Third is the output unit operation, "delivery of search results." The delivery might involve a set of bibliographical citations, a set of abstracts, or the source material themselves, depending on how the system is to respond.

These are the unit operations that I have used to organize my own thinking, and I'm asking you to accept these, for the purposes of the current discussion.

Now to the matter of centralization, decentralization, and specialization. I'd like to consider the topic from a historical point of view for a moment, to see how the centralization theme developed. I'm not going to bore you, certainly, with a discussion of the information explosion, and the reasons why people became concerned with nonconventional information retrieval systems, but rather I'd like to discuss some of the consequences of this information explosion that may have escaped your attention.

I'm told that professional folk many years ago were able to read the total literature in their field; and it turns out that when they were able to read the total literature of their field, the types of questions they asked of the libraries or of the information systems of those times, were based on a recall mechanism--in other words, having seen the material in the literature they were able to recall something about what they had seen, and therefore the systems that were developed to aid them in getting at their materials, weren't challenged as much as they might have been. This is by way of saying that a fairly shallow level of analysis was enough to provide reasonably adequate service to the user. If you'll contemplate for a moment what happens when one is not able to read everything in one's professional domain--the nature of one's question tends to change. It's no longer a matter of recall, but rather a matter of asking questions that relate to specification of a number of characteristics of a problem (a *problem* that one *has*), rather than *remembering* something that was in a document that one has read. If one were to specify characteristics of a problem, then the analysis required for materials that go into a file has to match the depth to which one wishes to penetrate the subject matter when a question is asked. For example, if a question is to be posed that has four or five characteristics to it, all of which must be present in a document that's provided as a response, and if other questions with other characteristics are to be answerable from the same document, the depth of analysis has to increase to the point where all of these characteristics are made explicit. We have seen information systems that have been developed that attempt to provide this depth of analysis by providing 10, 15, in some cases many more index entries, or descriptors.

When individuals attempt to perform this depth analysis, none has enough resources to do it, for all of the materials they consider to be of probable importance at some later time.

Thus, the centralization theme starts. One starts looking to a more central organization with the available resources to do the job adequately. Let us assume that this central organization is a government agency. Let's assume that this centralized activity starts somewhere in the Washington area, and they acquire their materials, they analyze the materials in some depth, and do all the things they should to get a system going. Then they say, "I'm now open for business, and now start feeding questions to me." The questions, after a long induction period, start flowing in.

After a time, it turns out that the centralized facility starts encountering problems. I'd like to characterize these problems as "queuing problems"; which have two identifiable dimensions. One is a physical queuing problem: people were accustomed to depending on human literature searchers to do their work for them (if they were fortunate enough to have some that they could use), and they were very often willing to wait months, or even longer, to have a thoroughly adequate job done. But when they approach and address themselves to a computer-based system that's supposed to operate quickly, days or weeks are too long a period to wait. So, a queuing problem of a physical nature develops as too many questions start coming in, and people become dissatisfied waiting in line for service.

The second aspect of the queuing problem is "intellectual." There is nothing in computer-based information retrieval systems that's different from the human systems of the past that causes it to be possible to ask questions without the system understanding what the question means; thus, a reference librarian function is required here, so that the question can be negotiated. The problem now emerges of accumulating enough people in a centralized location who have enough knowledge of the subject matter of the questions to be able to interpret them properly and deal with them in some adequate way. So, with not enough skilled people available in these central services, one begins to observe an intellectual queuing problem. In an attempt to provide service rapidly, to amortize the physical queuing problem, shortcuts on question negotation may be used, and questioners become dissatisfied.

The centralized facility begins to look for ways to solve their dilemma. One way has been to decentralize the activity and spread the grief around the country. When they start to consider decentralization procedures as have been done in a few systems, there is a tendency to say: "We will continue to perform all the input functions since we're organized reasonably well to do this now, and we will provide replicas of the searchable files, and perhaps even the documents themselves, to a number of organizations around the country, and let them worry with the problem in their region."

What happens? When the decentralized activities are in business, (generally in institutions where the requisite subject expertise is present) it is hoped that the intellectual queuing problem, and the physical queuing problem is solved because centers were chosen where adequate computer and related facilities are available to provide service quickly and efficiently. However, these centers do not operate totally without difficulties because they face economic questions. Also some of the physical queuing problems are indeed not being totally resolved because they find eventually that they must start "batching" their questions in order to achieve some economies of scale, so that some service delays begin to emerge. After some resolution of the queuing problem, you'll find another problem emerging. For example; a specialist in a given subject area says, "You know this was a beautiful system for somebody, but I don't really know who; it certainly isn't a solution to the problem of finding information in my specialty." You begin to get a pressure for the creation of a specialized information center in some fragment of a field. These things cost money. So, you go to that ultimate source of financial aid (or what used to be the ultimate source of aid), in Washington, and procure the dollars to set up the specialized information centers. This opens the flood gates and many such specialized centers eventually are scattered around the country, as in the case of the decentralized centers.

These centers start operating and knowledge of the availability of these centers begins to get around. Let's picture now the problem of an individual in Kalamazoo, Michigan, who wants an answer to a question, and he's told by the government agency that one particular specialized information center can handle his question. He approaches the center and finds that they're still in the formative period, and serving only their own people in the area; and he must "give up," but not without becoming annoyed. His annoyance is not decreased by knowledge that his taxpayers' dollars are being used; and second, he was annoyed anyway because he thought he should have had the opportunity to develop that particular specialized center. Finally he says, "Well, I want to have my questions answered, and I don't want to wait, and if this center can't serve me, let me have a copy of their tapes."

The government agency considers this demand for a while and eventually finds that they have to give him a copy of the tapes. Suddenly he has the tapes, and he has represented himself to have an adequate computer to be able to do the searches, and he's in business.

But, it doesn't take very long before he gets a little more ambitious and says, "Well this is only one specialized center, but my interests are interdisciplinary, really, and I would like to have the tapes from a number of specialized centers." Well, you don't have to extrapolate very far before you recognize that with the number of specialized information centers that are and will be available, a monstrous financial and traffic problem will emerge, with copies of tapes (updated periodically) going to Kalamazoo and several thousand other places around the country, in order to be able to rationalize some of these problems that emerge in servicing questioners. It is not merely the problem of having to pay for this, but you also develop a mishmash of criss-crossing efforts, with all the specialized centers not having the same indexing policy, the same thesaural policy, and certainly not the same policy for recording materials' index entries on tape. It's very difficult to merge these products. You find that the user ends up by dealing with these tapes like a library, in other words, "Put them on shelves, one next to the other, and exploit the tapes one at a time, in order to satisfy requirements." It is quite uneconomical and inefficient as well.

The next solution is the network theme, which says that we can now contemplate a mechanism that will provide data, voice, and image transmission capabilities, that will permit one to sit at one's desk and communicate in any mode that one wished from anywhere. With the data and voice communication capabilities one can negotiate a question with a distant file holder; with the image transmission capability one can start looking at some of the consequences of the search performed. So, let's look for a moment at what this network would look like.

Consider four nodes of a network, interconnected in every way possible. We can ignore the decentralized activities, because it is no longer necessary to copy tapes and make them available, and we are left with the centralized and specialized information activities that now can interrelate, or be tapped, by the network. The four nodes can be interconnected any way you wish. The centers link up with their nearest node, and presumably we're in business now to serve this gentleman from Kalamazoo, who now just needs a line into the network, which provides him with access to all the national resources available for exploitation.

Now the problem really starts, because once he taps into the network (and let's assume this Kalamazoo gentleman is an interdisciplinary type), and he has a question that cuts through many files that are "on the air" in this network, for example, if his question is in the field of biochemistry, then there might be responses in MEDLARS, in a number of the specialized centers of NINDB, *Biological Abstracts*, *Chemical Abstracts*, *Excerpta Medica*, and a few others as well. So he dials the network; someone has to answer his call--perhaps a telephone operator, who says "What's your question?" He recites his long complicated question in biochemistry, one that floors the operator who now must switch him to some appropriate point where his search might be fulfilled. What sort of telephone operator do we need in the network to permit an adequate switching job to be performed? The telephone operator, if we consider the telephone company analogy, might have a telephone directory with the names of all the services and all of their telephone numbers. That doesn't do much good, because it still doesn't suggest the best place to start looking in order to limit the cost of this search. So I wonder what level of locator tools have to be made available to this individual. You certainly don't want to have a set of specialists sitting in this telephone exchange, all of whom would get in the conference and decide where to look first. Here is an unresolved network problem that must be resolved before network planning may proceed.

Perhaps that's enough of a picture for this presentation, since it suggests how interesting and important is the issue that needs to be resolved: how to balance off the resources that are being allocated for specialized and centralized activities, decentralized activities, and efforts directed to network activities, in a period when one is trying to achieve service objectives rather than just waiting until something happens that will solve the problem.

So what I'd like to do now is to spend a few moments in talking about the way I run my life now, given these issues and problems as I see them. Fortunately I am in a university, with an opportunity to encourage good doctoral students to do their research in areas that seem fundamental, and that might throw some light on these areas. Let me tell you where we're directing our efforts, for better or for worse.

First, with regard to the analysis unit operation. Here I see a very interesting question which must be considered: what is the best surrogate that should be provided that will be useful to an enquirer of an information retrieval system? For 50 years or so, it's been assumed in various fields, like chemistry, that a good informative abstract is the best surrogate for an individual; you provide him with an abstract, and it will serve one or both of two functions. One is to serve as a substitute for reading the full document, and the second is to provide an indicator or predictor of the relevancy of the document, to permit him to make a decision as to whether to read it or not. It is in the second function that we have done some work, with regard to how good, as predictors of relevance, are the various surrogates, like citations, extracts, and abstracts of different types. I'm not going to suggest here what the results of this first investigation were, because it would take too long to give you the full story, and I think it's dangerous to just give you the conclusions without giving you the full story of the investigation.

A second area where we have tremendous interest is in the area of vocabulary control. Here we have two types of work going on. One is with regard to the systematic development of a thesaurus. Is it necessary, we ask ourselves, to develop thesauri, cross-reference systems, or vocabulary control mechanisms, in general, on an ad hoc basis, as they are developed today? We wondered whether a systematic look at how these terms fit into a hypothesized taxonomy, or framework of a given field, might lead to a thesaurus that might be used in an information retrieval system with predictable results. As an example, one student has taken this problem for his dissertation in special education. One other effort in the past was in paint technology, using such a systematic approach.

The second type of work, and this is another dissertation activity now going on, is looking at thesauri that have been developed in an attempt to develop a general methodology for testing and evaluating thesauri. How good are the thesauri in terms of the criterion one must use in an information retrieval system: how well does it improve the relevance of information provided to the individuals when a question is asked.

Does the use of a broader term provide a larger number of references than the narrower term that might have been used in the index? Does a narrower term provide a subset of the relevant material? We're trying to develop this general methodology for testing thesauri, using a given file, for which relevance information based on questions answered for motivated users, is available. We can now look at a thesaurus that has been superimposed on a system to overcome the inconsistencies of the indexing in past years, to see whether it indeed does the job it's supposed to.

We have one doctoral study comparing batch versus interactive systems from various points of view, not the least of which is an economic question. At what point do batch and interactive systems cross over with regard to economic and other related factors? Another dissertation is attempting to develop and test a conversational mode for teaching search strategy development (using a time-sharing computer).

In the area of acquisition of materials and storage of documents that have been acquired, we're interested in studying this matter from a resource allocation point of view. One can contemplate that one document can be kept in one place in the country, and everybody that wants it can order it through an interlibrary loan procedure. Or one can contemplate a network, providing image transmission capabilities, that permits distant viewing without physical removal of the document. Or one can go another route and say that microform copies of documents will be provided wherever required.

Is it better to do this, or is it better to stock a number of warehouses around the country or regional centers that will provide this service? The resource allocation problem here seems interesting, and is analogous to a pipeline or warehousing problem in the industrial engineering field.

We are also interested in total systems design, and here we have one student who is just getting started on the PERT/CPM approach to looking at systems design and operation.

When all else fails, we set up an experimental seminar, as we did last term with the advanced students in this area, to start looking at systems' design problems from a resource allocation point of view; and a computer game was developed in class that is operational now, ready to be tried with master's students some time during the next year, to see whether it is useful in teaching systems' design.

It is perhaps obvious from the foregoing that simplistic statements regarding centralization or decentralization are dangerous to make, since even small changes can change radically the ecology of the systems that mesh, in order to provide a useful information environment.

CONCERNS OF COMPATIBILITY AND ANALYSIS OF RESEARCH: ARTICULATION OF INFORMATION SYSTEMS

Ted Ward

Information retrieval promises to provide conceptual and technological tools adequate to meet the data and documentation needs of an increasingly verbal culture. Scientific inquiry today knows no bounds; the interface between biological and behavioral sciences poses especially complex issues. Not only is the problem that of lateral communication across fields in which data are collected and reported in very different manners, but there is a problem of vertical communication involving several levels of verbal sophistication. In earlier days it was possible to design documentation centers and information systems with only a researcher's uses and perceptions in mind; not so in today's behavioral science arena where the multitude of practitioners constitute a major communication concern for the information system designer.

THE DAWN (OF A NEW PROBLEM)

The advent of computer assistance in verbal information processing has been generally viewed as the dawn of a new day. It has been too easy to forget that information storage and retrieval is not new; the staggering problems which computers promise to solve and the glamor of computers themselves tend to obscure the fact that brave and lonesome men have been devoting themselves to library management, abstracting services, analysis centers, and document dissemination operations for quite some time. The new wave of computer applications specialists has tended to dramatize what *can* be done rather than what *should* be done to solve the problems associated with burgeoning documentation.

A VIEWPOINT

It is only fair to state the biases of this paper. It proposes optimization rather than polar choices. The drawing of careful balances between or among the contrasting probabilities which affect a design or plan is the essence of sound planning.

Thus, to view the design of information systems as *library versus computer* or *man versus machine* is unscientific and unproductive. The dreamer sees what could come about if all circumstances are immediately configurated like the dream; the realist sees the danger of abandoning the relative predictability of a present mode in favor of any unproven alternative. Optimizing requires a search for the intersection of the curves drawn by these two sets of values. The biases of the paper are activated by a dissatisfaction with present modes and an anxiety about the potential calamity involved in arguing for the wrong dream.

CENTERS AND NETWORKS

Wild enthusiasm for defining a model information system along the lines of what *could* be (rather than what *should* be) produces a predictable tendency to think big. Thinking big leads to plans which place the highest values on centralizing, networking, and high-speed communicating of conformed data.

Although it is a truism that information systems are effective to the extent that they meet real needs of real users, it is nonetheless easy to forget the user. What problems are created or solved for the user by centralization, networking, speed, and data format conformity? None of these is "good" *per se*. Does anyone seriously await the day when the scientist's notebook will be replaced by wireless two-way voice communication with a remote-accessed computer file, networked into all other scientists' files? Thinking big can divert us from seeking timely solutions for the functional here-and-now problems. Such solutions could help to unite scientists; to delay in the hope of realizing greater dreams can aggravate the already serious backlog of unsolved scientific communication problems.

Distinguishing *centers* from *networks* could be a useful semantic device. *Center* would be the term used to indicate the individual information processing unit designed to meet a specific set of known needs for a known understood user group. Matters of geography, field of research, level of sophistication, and habit (style) of information use are the principal variables which concern the designer of a center.

Network would then be used to indicate an articulated combination of component centers--the typical product of "thinking big." The major concerns of the design for the network are preservation of the essential characteristics of its component centers, optimization of network intercommunication, and appropriate transformation among levels of sophistication.

THE RIGHT TO MEET NEEDS

In the face of super-system dreams and designs the rights of the "little man" must be set forth. The little man--and these men are not really small in terms of their past or potential contributions to the world of information processing and dissemination--are those who operate the many extant information processing operations: libraries, abstracting services, document centers, and so forth. These men and their agencies not only have a right to survive, but they have the right, and the resources, to contribute substantially to the coordination and interlocking of information centers. They have a right to help structure the dreams of networks for the future; they are invaluable to plan the sorts of contributions present centers can best make. To overlook the experience and the aspirations of those who have made information centers the focus of their careers seems neither desirable nor tolerable. It could very well be that this oversight has been one cause for the demise of several of the abortive computerized information retrieval systems.

Those who are responsible for the maintenance of operating centers can serve a valuable purpose in keeping the dreams reasonably well tuned to reality. There are needs for information which are already known and understood; there are user habits already defined and accommodated. The on-going information center has a right to continue meeting these needs. Planning of new systems or networks must account for continuance of services to established users.

UNIFORMITY WITHOUT CONFORMITY

Interlocked information centers require certain sorts of uniformity. There are two essentially different ways to approach the problem of uniformity: (1) to begin with a model of the ultimate network and work back toward the component centers or (2) to begin with the descriptions of the component and potential component centers and to design toward a network. The first approach requires that constraints be imposed upon any pre-existing components of the network; these constraints will result in conforming the component centers to the design characteristic of the network. The second approach seems preferable when pre-existing centers are to become part of a network, because it treats articulation as a secondary concern, allowing primary attention to remain focused on the several sets of users and their needs.

Monolithic conformity has not yet been shown to be an absolute requirement for articulated information centers. True, mandated conformity does offer a tempting alternative to the predictably tedious and difficult design problems common to the planning of uniform articulation among autonomous centers. The price of this tempting alternative is insidiously high; the required administrative bureaucracy leads to increased remoteness from the problems of users; which, in turn, leads to hopeless communication breakdowns.

Individual centers can be designed or redesigned to fit into an articulated network by accommodating these minimal guidelines:

1. Minimize dependency on peculiar hardware;
2. Minimize coding (maintain high level of meaning in the data entries and in the file-search descriptors);
3. Develop standardized or compatible languages of indexing (and/or design transforming buffer functions for the computer to make file-to-file language transformations);
4. Develop standardized or compatible formats for the reduction of data through abstracting;
5. Design administrative procedures and communication structures to differentiate the labors of each center (to minimize duplicated effort in favor of a maximal breadth of coverage);
6. Develop computer programs for format transformations and thesaurus translation to provide center-to-center compatibility.

CHALLENGING THE ASSUMPTIONS

At the risk of setting up straw men, the following list sets forth assumptions which seem to be implicit in some of the large-scale information networks being developed today:

1. Merely setting up an information network will bring potential information users into productive contact with important information which otherwise would not be available to them.

Comment: For scientists this is true only of peripheral information, for practitioners it is not true at all. Scientists have highly developed but largely informal systems for staying in contact with the central core of data which concerns their work. Practitioners need more than additional information systems--they need training in how (and why) to use outside information.

2. The existence (availability) of an information center or network will generate new habits of information use.

Comment: While this may contain borderline truth, it falls into the realm of wishful thinking. Either an information center makes contact with users through their established habits (minimally modified to accommodate procedural requirements) or else the center must engage in extensive training programs to develop the new habits required of users. It would be useful to test the hypothesis that there is no predictable relationship between the existence or non-existence of a formal information system and the information-use behaviors of any professional group. (This hypothesis does not contend that there is no relationship with the information-*seeking* behaviors.)

3. Effective nation-wide information networks can be designed *de novo*, while avoiding competition with existing information centers (abstracting services, microfilm, libraries and document dissemination centers).

Comment: The key word is "effective." It seems illogical to try to solve the problems of multiplication of information sources through the creation of one more source--unless the newly created source can eliminate the need for two or more others.

4. A network designed primarily for researchers cannot be useful for practitioners.

Comment: Experience to date supports this assumption, but computer classification and transformation of data may someday demonstrate that one network can meet the needs of several levels of information users. Meanwhile it would seem advisable not to sell several audiences on a system which will meet the needs of only one of those audiences.

5. Studying the user is a once-only job.

Comment: We can never know too much about the needs and characteristics of the clients of an information center. Effective use of the center is a function not only of design characteristics of a technological sort, but it is a function of the degree to which these design characteristics accommodate the human characteristics of users. Today users "jump lines" more easily and frequently; in fact, a major motivation for information centers and networks is the growth of interdisciplinary scholarship. Continuous study and systematic in-system feedback on user satisfaction/frustration must be built into any sound plan.

6. There are vast quantities of information which never become available to potential users.

Comments: This assumption is clearly a statement of fact. The issue is whether or not this state of affairs should be changed. Researchers are not contending that this "fugitive" literature is needed in their work; they are, instead, inclined to trust professional reviewing and abstracting services, slow as they are, to serve as a sort of filter--sparing them from wasting time with inconsequential material. It is said, especially in education, that practitioners need access to this fugitive documentation. One wonders why, since it is apparent that teachers and school administrators are unsystematic users of professional literature--possibly a function of the disorganized and massive sprawl which is characteristic of professional and semi-professional literature in education. Thus when ERIC or any other network proposes to process "fugitive" documents and materials the issue is not so much whether the machines are capable of handling this mass, but whether the human processors will be able and willing (enough time? appropriate effort? discrimination skills?) to handle this quantity of disparate pieces.

7. What *can* be done *should* be done.

Comment: This is the great technological trap of our era. Technical capability teases us into building elaborate systems to accomplish missions of dubious worth. Research and development in information systems should be accompanied by sober philosophical examination of values and relative merits of different approaches to problems. Technological capabilities raise difficult value questions since human behavior is today responsive to and reconstructed by the time-effort-skills demands of the technology.

RELATING COMPUTERS TO INFORMATION PROBLEMS

We still hear well-meaning people say, "Yes, we have the computers, but we lack the software." I'm weary of this. It's neither hardware nor software that holds us back. What we lack are hard heads. We should stop arguing about characteristics of Utopian systems, and get about the business of applying information retrieval procedures to known needs in reference to known users.

Since we believe this at the Learning Systems Institute, our studies have proceeded through these steps:

1. Develop a view of the role of information use in the professional areas with which we are concerned;

2. Build a general system of computer programs capable of assisting in the various sorts of information use in the professional areas;

3. Generate applications of these programs for various information system configurations;

4. Field-test and refine procedures;

5. Evaluate the needs met as well as the needs not met by each configuration;

6. Redesign programs, applications, and configurations in terms of the feedback from the evaluation.

System design criteria which we have imposed on our own information system projects are presented here as a summary, not because they are right but because they reflect the best thinking our group of scientists and engineers has generated from careful study and analysis of contemporary needs.

What comes off the computer in response to a search must be verbally meaningful (not just a list of code words or serial numbers).

On-line search operations must be available and must permit the inquirer to use his own natural language.

Abstracts of documents or data files must be available directly from the computer (to allow the inquirer to scan--or even study--as he participates with the computer in the narrowing down of his search question).

Use the computer to solve problems, not to make them worse.

Thus, before you begin to design, *make sure you understand the problems*!

RESPONSE OF A SYSTEM OPERATOR

Claude Marks

When I was asked to represent the University of Texas Special Education Instructional Materials Center, as a participant in this conference, I thought I knew just exactly what was meant by "the response of systems operators." As time went by, I began to feel less and less confident. I do know that I have disproved Wendell Johnson's First Law, namely, the infallibility of the labeling process. Johnson postulated that the labeling process was a creator in and of itself, and once one called an object or a person by a particular name, that object or person became that thing, whether in fact, it or he was that thing or person or not, did not matter.

I have been called a "systems operator." I wanted, but did not become, that. Rather, I am part of a highly energetic, particularly peripatetic organization of professionals, who function most of the time as systems operators. I am not at all sure that this organization is, in itself, a "systems operator."

I am beginning as a noninformative, noninformation scientist, to believe that the "system" might be operating us. Today, I would like to present information for your careful processing about the use of the computer in the information retrieval program at the University of Texas Special Education Instructional Material Center (which I shall now call the SEIMC). First, I will present a brief history of the development of the SEIMC network from a concept to a quasi-reality. Second, I will describe the operation of the UT/SEIMC with an emphasis on the information retrieval program. Third, I would like to describe to you, as best I can, what I call the "nitty-gritty" aspects of the information retrieval program, and lastly, present for your consideration, and as information scientists, the real problems that we have encountered.

The SEIMC network had its birth in the findings of a presidential task force in 1962. This group had toured European countries to learn of advancements by European special educators. One of their findings was the existence of excellent instructional material resource centers for teachers of handicapped children. The objective of these centers was to provide ready access to a wide variety of instructional material and to information related to the education of the handicapped. In 1964, this concept of an instructional material center was implemented in the United States. Dr. James McCarthy of the University of Wisconsin, and Dr. William Read, of the University of Southern California, were encouraged to implement pilot materials centers by funding from the research branch of the Handicapped Children and Youth Division in the U.S. Office of Education, which is as you may know now the Bureau of the Education for the Handicapped.

The childhood phase in the development of the SEIMC network was entered in 1965 at Atlanta, Georgia. The first two pilot centers experienced such fine early success, and attracted such attention, that the Southern Region Education Board initiated efforts to expand the number of such centers. Long an efficient vehicle for the articulation of innovative programs, the SREB invited Dr. McCarthy and Dr. Moss from the Handicapped Children and Youth Division, and representatives from three major Southern universities who had expressed interest in the materials center concept, to a meeting in Atlanta. The original intent was to seek funding for materials centers at the University of Kentucky, the University of South Florida (Tampa), and the University of Texas. During ensuing discussions, the network concept was generated. The three university representatives had three unique contributions to the discussion, and eventually these ideas were reflected in their proposals, and eventually embodied, as essentials in the network concept. These were (1) the development of materials, (2) the evaluation of materials, and (3) computerized information retrieval. Dr. William Wolf, who was Chairman of the Department of Special Education at Texas, envisioned these three centers, plus the two pilot centers, tied together and sharing a common computer for the rapid and efficient retrieval of information pertaining to the materials and documents which were acquired.

Other special educators around the country became infected with the early enthusiasm, and ERIC saw how such an expansion of the centers into a network would be a viable dissemination unit for a clearinghouse. The Office of Education funded eight new centers, bringing the total to ten, each center responsible for service to several states. The following year, the U.S. Office of Education funded four additional materials centers. At this point in time, these fourteen centers have developed within their own regions more than 70 associate or affiliate centers, in order to expedite the dissemination of information.

Although the fourteen centers were bound by common objectives, each in its own way demanded some justifiable autonomous objectives. Each center stated these objectives within a framework of the two common or generic objectives: service, and research and development. At the University of Texas, we have stated these objectives as follows: (1) To serve as a central depository for curricula and instruction materials for exceptional children. (2) To receive, distribute, and loan curriculum guides, research reports, programmed learning materials, films, film strips, and et cetera. (3) To provide an index system on subject matter areas relating to the education of exceptional children. (4) To provide local, state, and regional consultation and guidance services relating to instructional materials. (5) To offer credit and noncredit short courses, institutes, workshops, for the purpose of acquainting participants with new materials and methods. (6) To evaluate, adopt, and modify existing instructional materials, and to create, assess, or assist others in creating and assessing, instructional materials for exceptional children.

The directors of the SEIMC's relate within a formalized association, which has elected officers, and all other official accoutrements usually associated with associations, including committees. During one of our early meetings, we formed a computer committee, whose function was to address itself to the business of information retrieval. At that time, there were four centers which expressed interest in information retrieval, and each of these centers has a different approach to the problem. The system at the University of Texas took its initial form under the consultation of Dr. Al Vale, then of the Linguistics Research Center, and now Professor of Computer Science and Library Science. Dr. Vale had developed, with his staff, a coordinate search program, which we adopted with the aid of a student.

In the beginning, we had a choice of direction in terms of what body of data would be stored on the computer for retrieval; that is, the entire abstract of about 250 words from a document plus descriptor terms, and bibliographic information; or the descriptor terms plus a document number and location number in our holdings. There were two major reasons for selecting the latter route. One, we did not want to go against the advice of our consultants, and get involved in machine indexing; and two, the logistics of preparing the data and entire abstracts from machine storage were staggering, if one projected the job over years the number of documents. As our system took form, it included the following components: (1) a CDC 6600 computer, with peripheral processing equipment, on campus, approximately a mile from the center; (2) coordinate retrieval programming; (3) 200 units of central disc space, plus the tapes; (4) respond system for on-line time sharing; (5) a Teletype Corporation Model 35 Teletyper at the materials center; (6) a Recordak microfilm reader/printer; (7) microfilm cartridges for each one of the files: one on ERIC documents, one on professional books, and one on instructional materials.

The following is an outline of the procedure of processing a request for information in this system:

First, the user identifies himself, and provides the number of the account to which time used on the remote station is to be charged, on a Model 35 Teletype console; it is attached to the Control Data 6600 computer.

Second, the user types his request in natural English, observing only a few rules designed to insure clarity. Basically, any statement of his need, which is clear to a librarian, should suffice for the computer. The user then determines the correct terms or phrases to fit his needs from the SEIMC thesauri; there are three of them: of descriptors for ERIC, professional books, or instructional materials. The request is then matched against documents in the reference collection, using efficient and accurate search stratagem. Those documents relevant to the request are selected by the computer.

Third, within a time interval measured in seconds, an answer is provided. This answer will be in the form of a microfilm frame number, and the materials center library classification number. The latter form will designate those locations within the instructional materials center network which have the selected documents. If the document has been checked out, the user will be informed of the return date by the library. If the user wishes to know what terms are available for making a request, he may consult the thesauri; he may wish to do this before the initial document search, or after the results of the search have been displayed. The thesauri will be stored on the computer as part of the data files. It will enable the user to become more familiar with the retrieval system, through interaction.

Finally, the user leaves the terminal to evaluate his output. He evaluates his output by consulting the microfilm reader/printer, searching for the book, journal, curriculum guide, or instructional material in the library, using the classification numbers.

The following information will be given on each microfilm frame: the microfilm number, the author, the title, publisher, subject area, the type of endeavor, the number of pages, the location, the indexing terms, and an abstract. This microfilm reader is right next to the terminal, and one merely swivels his chair to use it. Hopefully, the user will be satisfied by the results of his search; the search strategy should be designed to avoid search failures. Partial dissatisfaction, however, may arise from one of three causes: (1) the computer has retrieved documents that are not relevant to the question as stated; (2) the computer has retrieved documents that are relevant to the question as stated, but because of ambiguities or a misphrasing, the retrieved documents are not relevant to what the user actually intended in his request; (3) the computer has retrieved documents that are relevant to the question as intended and stated, but after consideration of these documents, the user has decided that he asked for the wrong topic.

In the first case, the problem is basically a flaw in the computer system; the librarian or systems analyst should be notified that this occurs. In the other two cases, complete rewriting of the question is obviously necessary.

In December of 1966, Allen Kent of the Knowledge Availability Systems Center at the University of Pittsburgh, consulted with the staff at the SEIMC about the general problems of (1) developing procedures for instructing the people served by the center in how to submit requests for information; and (2) abstracting and indexing of the material so as to make the information readily available in a relatively short interval of time. Professor Kent guided the discussions, and aided the staff in reaching some tentative decisions, procedures for abstracting both the instructional and the professional materials. Since the SEIMC gives priority to serving teachers in the field, it has decided to work with the available personnel to make these materials, instructional materials, available on the computer at the earliest possible date. The form for abstracting instructional materials was also developed during this meeting. The format was designed to facilitate a description and an analysis of the materials in terms of the type of material, the cognitive process thought to be involved, the content area covered, the disability group and size of the group for which the teacher requested the material, the level of instruction at which it would be applicable, and the possible use in counseling and guidance services were that the case. This format differed from the attempts made before Dr. Kent's visit; and a narrative abstract was no longer thought to be the most useful means of supplying services to teachers. Instead, the information provided by the instruction form would be interpreted by materials specialists, as to whether the document retrieved by the computer system was truly relevant to the request submitted by a teacher. The abstracting and indexing of professional materials, comprising over 5000 titles, was considered in terms of the amount of personnel available, the time required to write a narrative abstract and index each document, and the period of time in which the service was to be partially operated. Dr. Kent suggested that when time and personnel are limited, the table of contents might be used as a reliable abstract, together with careful indexing and a selection of descriptor terms.

A narrative abstract could be added to this initial phase once the funds and time were available. With this type of abstract, many of the materials in the SEIMC library would be available for machine retrieval at the beginning of Fall, 1967. Indeed, we did both abstract and include the table of contents, as well as all the bibliographic information. The fall of 1967 marked intensive efforts on our part to make our system compatible with that of ERIC, including the development of the thesaurus. For consultation here, we called upon, among others, F. Brenner from the American Petroleum Institute. The details of the tribulations here are beyond the scope of this presentation. Suffice it to say that we encountered some problems. Our system allowed for more flexibility in format in the thesaurus than did the ERIC system, that is, hyphenated versus nonhyphenated terms to eliminate some false drops; and indeed there was a period of six months or so when it was doubtful whether ERIC would encourage the inclusion of descriptor terms for instructional materials in the thesaurus. We still are not clear on this particular point, whether or not ERIC is going to include instructional materials terms. I think we're getting close to it. If they don't, I believe that they're violating one of their major purposes, that is, to serve teachers with the tools of their own trade.

Early in 1968, at a meeting of the directors of the SEIMC network, things began to mature. We were encouraged to continue to develop the instructional materials thesaurus, and to join ERIC and CEC in the development of a manual for abstracting and indexing instructional materials. In July of 1968 a training session was held at the University of Texas SEIMC for the personnel from all the centers, including ERIC/CEC; the purpose of this meeting will be to train and teach abstractors and indexers. In this way, the network will have at least one person in each center who has learned a set of principles for indexing and abstracting which will be common. This delay in developing a common system has not overly retarded our progress. It, along with the difficulty in locating a sufficient number of trained abstractors, during the fall and spring, has slowed us down. Perhaps, though, the slowdown has had its assets; for example, our staff, if not pleased with the quantity of abstracts, thinks that through this period of reflection, we have gained some quality.

In 1968, the SEIMC employed 15 graduate students in special education of teachers, to tackle the job of preparing over 5000 abstracts on instructional materials. We've experienced no major machine or programming problems. What can be done far surpasses what is being done, basically, for lack of sufficient number of qualified people to abstract. We have attempted to use people from outside the special education profession as abstractors, but for the most part it has not been successful. We found two who are competent as abstractors of professional texts, but not as abstractors of instructional materials.

The SEIMC computer retrieval system provides for storage and dissemination of information for all professionals in the field of special education. We feel that our system has for its core an aggressive and efficient retrieval program. Some difficulties have been encountered by users of the system, as well as some specific to our particular system; that is, some of the things that I'm going to relate to you would be general to users of any system, and some are unique to ours.

Computer scientists have not provided simplified explanations for the user. When the central computer source provides a user manual, the instructions must be concise and explicit, since the unsophisticated user is probably not familiar with information science terminology. Although the respond system is operational, frequent technical improvements, which upgrade the system and make it available to more users, leads to system malfunction during the improvement process. The resulting inconvenience to the user is often interpreted as a systems failure. Frequent and nontechnical consultation could help alleviate this problem.

Ideally, the central computer source should be in close proximity to all terminals, and provide communication in natural language. Difficulties with a program's adjunctive materials have also appeared. Adjunctive materials include, for our purposes, the thesaurus, the data bank, and the Teletype procedure. I will discuss each in terms of the problems encountered.

The materials I am discussing now were preapred by a librarian and a curriculum consultant. Educators use our system most, except for the ERIC file, which is widely used by those in social work and sociology.

Our thesaurus is a compilation of terms from many fields: medicine, education, psychology, etc. Professional decisions are required on the selecting of terms, the matching of equivalent terms, the recognition of nonequivalent terms, the relating of similar terms, and the adding of cross-references. The user must be familiar with the terms included in the thesaurus, as well as the terms of several fields. Unless the cross-referencing is extensive, problems arise in the interpretation of requests. The addition of terms to the thesaurus is extremely difficult. Research in the contributory professional fields is continuous, and the terminology of each field is subject to change. Often, reinterpretation of key terms results. Others have pointed out these same difficulties.

A specific amount of disc space on the central computer is assigned to each user. As the data bank is enlarged, the file storage space is reduced, thus eliminating free space for computation purposes. The data bank originally consisted of 18 disc spaces, 16 for professional literature, and 130 for ERIC documents. With all the files loaded, only 36 spaces are free for computation. Recently, the instructional materials file was increased to 26, thus reducing computation space to ten; and we have all three files on the disc. Each request averages three disc spaces, therefore permitting a maximum of only two requests.

When traffic on the system is heavy, turn-around time is lengthened. The system is constructed to provide a status report on requests. However, some requests never advance further than the first status reports, which indicates presence in the queue. Some users must wait as long as 30 minutes for their request to be processed. Some searches cannot be completed due to "down time." Recently, because of breakdown of circuitry on the machine, an engineer shut down the machine for minor repairs, and he neglected to turn off the 400 cycle circuit for operation of the machine, and instead turned off the air conditioning cycle. It burned out about a quarter of a million dollar's worth of circuitry.

Tape reel numbers are sometimes reversed, resulting in a delay in procedure and unusable response. Logical connectives, "ands," or "or's," frequently are rejected, or misinterpreted in the search. Sometimes, it is impossible for us to use more than three descriptors. It is possible for us to use ten at this particular time. Sometimes, the systems operators in the computer center just won't let more than three get through.

An example of misinterpretation in a search occurs when a term "primer" has to be retrieved, and the search pulls the term "preprimer." The situation illustrates a contextual, rather than a structural, search. An asterisk is used to denote descriptive terms in the documents which emphasize the entire subject. Some searches disregard the asterisk, and the result is that inaccurate and bunch of garbage.

Not wanting to end on a pessimistic or negative tone, let me add some information here about a system which is nearly operational and will help solve most of our problems in a man-machine interaction. In order to alleviate these obstacles to the individual user, the computational center, the linguistics center, and the SEIMC have been developing a general-purpose system for managing data in a time-sharing environment. The SEIMC is monitoring the development for use in document retrieval. The management system known as RFMS (Remote File Management System) will enable individuals and organizations with large and complex data files to manage their files with speed and ease, which is expected of a computer of the magnitude of the 6600. RFMS is oriented to the nonprogrammer user, who after learning some basic commands for communication, can work directly with the computer to manage his data base. These commands are real-language commands. The system takes advantage of one of the most promising advances in information processing technology: time sharing, enabling many persons to use the computer simultaneously, thus reducing the cost of direct on-line access. RFMS permits the user to describe entries in the data base, to load them into the machine, and ask questions about them, perform calculations on them, have the data displayed on the cathode-ray tube, to obtain hard-copy reports, and to update and maintain the data base. It is designed for use on the 6600. RFMS represents a new, integrated approach to the entire range of data management problems. Some of the major design features are: (1) RFMS makes the data management capabilities of a power machine directly available to the user without the necessity of communicating his requirement through an intermediary programmer; (2) RFMS is conversational: that is, during the man-machine dialogue, the system may ask the user to supply parameters, control information, file name, operations to perform, format desired--while the user, in turn, may ask the system to define a term, comment on a process he or she does not understand, tell him what steps of a procedure are available, explain error messages, or give the user other tutorial help; (3) RFMS provides the user with complete, integrated set of generalized set of data management's tools, including capabilities for manipulating data files, in generating reports for graphic displays, performing on-line updating of the files and retrieval, and managing many different kinds of data bases; (4) The RFMS file organization employs lists, tables, and inverted tree data dispensers which make optimum use of random access mass storage devices. This permits the rapid retrieval of isolated entries in the file, while maintaining the hierarchical relationships among the various data items; (5) RFMS operates under a system that allows many users to have simultaneous time-shared access to the computer. This makes maximum use of the machine resources, reduces turn-around time, particularly for fact-retrieval problems, and permits the sharing of data programs and the sharing of costs among a large number of users.

A VALEDICTORY

Kenneth R. Ingham

I would like to tell you a little more about my background in order to explain some of the motivation for my coming here or feeling that this conference would be interesting to me. My principal work at MIT was on the MIT reading machine system. Our group, for example, a year or two ago did the initial hardware construction and software design of the publisher's tapes or Monotype tapes to braille translation system. Since then we have been working on a complete reading machine system which now accepts as input inkprint materials, and translates the text to speech or braille with an error rate of less than one error per thousand characters. With this sort of equipment now available, both in our location and in other places around the country, a growing volume of machine-stored text now exists.

With these developments in mind, it is clear that some old problems will be made more acute and some new ones will arise. Allow me to group these problems, as I see them, into three main categories.

First, with a growing volume of machine-translatable text on our hands, we should devise the proper and most economical means of storing this material. Here we are concerned with storing not only citations but whole texts as well. It would be a grave error, as far as the visually handicapped are concerned, to destroy this material, as the existence of text in machine-translatable form would provide the first opportunity to meet the reading needs of the blind. At present not only is it difficult for a blind person to determine whether or not a book exists in braille or speech, but he will inevitably have a long wait before it is in his hands. This delay is prohibitively long if the reader has to wait for the book to be embossed into braille. If he had a quick means of determining the existence of a given book which would be machine-translatable, the braille copy could be produced and delivered in a few days' time. The fact that books with a low demand usually are laboriously recopied is a disgrace in view of the new equipment now available.

Second, as a researcher in the field of sensory aids, I am often forced to examine the current literature on particular topics related to blindness. Furthermore, the research is so diverse that I might find myself looking into fields that one would not ordinarily associate with sensory aids. I might have questions about psychology one day, and about physics the next. In any case, I must know what my colleagues throughout the field are doing.

Third, the average blind person has, as a matter of course, a whole variety of questions he would like to ask. If he is newly blinded, he will want to know about his condition, the medical aspects, the services that might be available, what kinds of aids there are. Here the blind person, as well as any sighted person with similar questions, needs to be instructed. The availability of knowledge in raw text form militates toward the organization of this knowledge to permit examination by, and the instruction of, individuals of categories two and three.

Thus it would appear that the efficient storage of text, the availability of knowledge, and the potential to develop instructional systems utilizing this knowledge represent three areas where problems exist, and, as this conference has indicated, the means to solution is near.

I will now try to relate the problems that I brought with me to this conference, and which I have just outlined, to the content of the conference. I hope that along the way I will find at least partial solutions to these difficulties.

At the first two sessions of this conference, the existing library systems were described. It was pointed out that while some of these systems, such as MEDLARS, have been around for a long time, their design has been motivated from within rather than as a direct response to user requirements. Furthermore, even though some of these systems are being automated, the speed at which the user has his request filled is not by any means ideal. It does appear, however, that a means to store and catalogue the texts now available does exist, and if the user can ask the right questions, any of the existing library systems such as ERIC/CEC could maintain the special materials needed by both blind persons and researchers in the field of blindness. The key issue seems to be "How do you teach the user to ask the right questions?" How do you interface the borrower to a system which, from many points of view, including cost, should be as efficient as possible?

The one ray of hope that indicates a method of attack to this problem was uncovered by the description of the Vision Information Center by Dr. Reinecke. He pointed out that the document retrieval system being developed at Harvard is being meshed with a computer-assisted instruction mode which permits the uninitiated to obtain instruction in the basic medicine of the eye. This means that a user whose needs are poorly defined in his own mind can acquaint himself with the thesaurus of the field of interest and in a manner which at the same time instructs him in some of the basic relationships of ophthalmology.

If a variety of highly specialized systems of this sort were available, the researcher would be able to explore the state of the art in a given field as well as achieve an interdisciplinary mobility impossible in a non-interactive system. Thus, I am suggesting that the document or text retrieval aspect of the International Research Information Service be put into the hands of ERIC/CEC and that the remaining educational and instructional aspects be implemented on a computer system which would be accessible to researchers and workers in the field of sensory aids and/or blindness.

Such a system should be highly interactive and utilize some of the new display equipments now available for computer output. It is also suggested that the major laboratories doing the research in the field be permitted a "key" with which to unlock the system's files and permit signed updating. Such a system does not exist, not even in people's minds; and therefore, the interfacing problem remains a challenging, but I believe very essential, part of the research that has to be carried out in sensory aids.

What I am referring to above concerning the highly specialized interactive system is not restricted to sensory aids. The concept is applicable in any defined field of research. The thesauri which, through the signed updating procedure, would be gradually built up and perfected by the researcher could be tagged to permit the relating of these vocabularies in the various disciplines. By cross-relating the sense of words read in several vocabularies, research could be carried on to improve the large batch-process system which might back up the smaller instruction ones. In any case, the user exploring the discipline, either for the first time or in depth to determine any new results, would have a far better chance of success with batch-process retrieval of the full text stored by the centralized system.

Finally, while the above approach would maximize the efficiency of the information retrieval systems available and proposed in the foreseeable future, the kind of "public" information retrieval problems described in the third category above would remain. Whereas the specialized systems described would be highly esoteric, the system required by the average non-technically oriented user would of necessity be much broader in content and a good deal more shallow. While a system to fulfill this need should be researched, it is true that the problems of the public are basically human ones. Perhaps the social worker or researcher who deals with the questions raised by the public and who might have the knowledge available through the specialized systems would be the best source for the answers. With the exception of students and researchers, there is no need for the general public to learn to use a machine system, and they would undoubtedly be better served by dealing directly with warm interested human beings. However, the random information about organizations for the blind and their services should be conveniently structured and made easily accessible to at least the worker who wishes to serve the public, and the best hope of achieving this would be some sort of computerized data management system. Such systems as described throughout should make full use of telephone services to permit access by as many potential subscribers as possible.

DISCUSSION SUMMARY

RESEARCH AND DOCUMENTATION CONFERENCE

Leslie L. Clark

With such a rich and varied set of papers as those presented in this conference, it might be expected that our discussion of optimum documentation procedures in the special field of sensory processes would reflect the general concerns of documentalists. It did. We attempted to constrain the discussion domain by posing the question: What advice could a "big" system operator give to those about to undertake the establishment of a specialized information center? The art and science of documentation, at its current stage, still prompted a wide variety of reactions to this question--reactions which reflected most of the major concerns of the information sciences. Nevertheless, the question hung in the air, like a probability domain, and it was gratifying to note that almost all of the discussion fell within that domain.

SYSTEM DESIGN: MACHINE FACTORS

It seems clear from what was said that the major problems, now as in years past, lie in the human factors associated with machine systems, and not primarily in machine systems themselves. We have had experience in this country with large communication networks, like the telephone networks; that experience indicates that future information networks may work.

They can work well to the extent that most users are satisfied most of the time when a link is established from an inquiry origin and that remote point at which an appropriate response can be made--just like the telephone network. They will work poorly to the extent that the information content is not taken into account. Telephone systems work because they provide redundancy, and because they are not concerned with the content of messages transmitted. The switching and transfer problems inherent in the operation of large interacting information networks will not be solved as easily, therefore, and will bear some careful attention now, before implementation, to avoid disappointment in their operation.

There also seems to be consensus that interactive modes will become the rule not the exception in information systems. In ten years we should see an increasing use of the technique of successive approximations to the satisfaction of a user's inquiry, through a kind of dialogue between the user at the console of a large time-shared computer and the stored data base.

Finally, there seems to be a consensus that the degree of complexity with which these systems will operate will be determined largely by the cost of the system and its operation. We could provide right now for such things as multiple audiences or disparate populations of users; for levels of privacy of stored data to which users have access; and for utilizing locally the results of an information inquiry in another system at a remote location.

Aside from the major factor of cost, then, this simple listing virtually exhausts the purely machine-oriented concerns in our discussion.

SYSTEM DESIGN: HUMAN FACTORS

In comparison, the influence of human factors on design and operation of machine systems loomed very large indeed in our concerns. Indeed, one might almost say that documentalists have become among the most expert in, and have the greatest sensitivity to, the human factors which operate at the man/machine interface. The inference from our discussion is that these factors pose the most frequently found difficulties in the design of systems, and in their operation, extension, and growth.

THE QUESTION OF RAW DATA

Let me dispose first of one item which serves to bridge the interface between man and machine, that is, the matter of incorporation of raw data into a "read/write/erase" capability of a large information network. I shall use an example which is conveniently at hand: the activity of the psychologist. Similar examples could be drawn from almost any area of basic experimental research. Even today, it is known that some psychologists would rather replicate an experiment than depend on the results of other psychologists' experiments. The reason is simple: there is a very profound identification of the psychologist and his knowledge with "his" experiment. His self-image, his legitimation as a scientist, his status in his peer group, all depend on this identification. It is not surprising, then, that a considerable emotional burden is borne by the relationship between the psychologist and his experiment. His feelings are often expressed in informal slogans, as for example, "He who would use another man's data would also use his toothbrush." Sociologists have analyzed this relationship, showing the context within which the researcher organizes his activity and the informal relationships which obtain. More generally, the emotional burden involved is a potent element in the facilitation or impediment of research; interested readers are referred to *The Double Helix* for a fascinating example of what I mean. Note, however, that the problem is enough recognized that provision has been made in some systems for differential access to a central data base. The method is simply to allow some users the right to read only, some to read and write, and the rest to read, write, and erase the data in central store.

I find it fascinating to speculate on the implication of Robert Rosenthal's work at Harvard, which indicates the profound influence of the working hypotheses of the researcher on the data generated in the course of his experiment--unless he is isolated from the lowest level of data gathering involved. To paraphrase his conclusions in terms of the above context, our psychologist's slogan may have to be rewritten one day thus: "He who gathers his own data may also force others to use his own toothbrush." Indeed, one inference that may be made with considerable justice is that in the future respectable researchers may regard automatic data collection (in "automated experiments") to be superior to the time-honored practice of the researcher to work directly with subjects in his own laboratory. There may well be then a volta-face on the issue of input of raw data into large information systems; and the implications for the definition of the user population, the machine configurations required, the costs involved, and the whole question of special-purpose vs. general-purpose information systems may all have to be rethought.

THE DEFINITION OF THE USER POPULATION

It is by now obvious that the decisions of the documentalist about the configuration of a particular information system are crucial to costs and the level of effort required to satisfy the needs of those who seek to find information. Since, indeed, society may have to pay for his decisions, now or in the future, the purpose of paper design and of "dry runs" assumes greater importance. One of the most fundamental decisions the documentalist has to make is what population he is designing his system to satisfy, and in terms of what profile of needs this population has. Some systems, such as MEDLARS, have the advantage of clear historical grounds of expressed need for information. Although there are difficulties enough in designing such powerful systems, the real difficulty comes in making them useful systems. Moreover, the question of the user profile must be asked before the system is designed, not after.

One way to deal with the problem is to build up a user profile based on a market survey. This seems particularly appropriate when no historical markers are available to indicate user needs. Another appropriate response is to provide in the system bibliographic references and syntheses of the literature by experts, rather than raw data or undigested research reports. There are some difficulties here, but this procedure seems to work well for the population of nonspecialists. It has the further virtue of satisfying least cost criteria. This is in effect the approach that ERIC has taken.

But the profile question does not evaporate when one opts for a bibliographic emphasis for the nonspecialist population; it merely submerges for a while. Once users have acquired familiarity with the system, and begin to use it, the problem resurfaces. The tendency, indeed, is for the development of several populations of specialists, each with different needs, within a "general purpose" population. And of course there exist small pockets of special interest groups even within a population of nonspecialists. These persons try to push the development of an information system in the direction of maximizing the satisfaction of their needs; moreover, they are listened to carefully, because they can express their needs explicitly. Such would be the case in the field of sensory research, and especially in the case of research related to blindness.

What can be done to satisfy the needs of special populations within the context of a general-purpose information system? If experience is any guide, it is entirely possible that the most efficient solution is to design separate systems for subpopulations of users whose needs are so disparate that they introduce conflicting requirements for system design.

It must also be kept in mind that many systems are designed by documentalists on the basis of market data which may have serious limitations. These data will not reveal the problems of information retrieval in a given field; they deal rather with the perception of persons in that field of what the problems are--not with what things ought to be, not with what they "really" are, but what the practitioner thinks they are. We need to ask what are the internally perceived needs of the people who really use the information systems we can build. In medicine, for example, it was long thought that common medical records might be "a good thing" for medicine; but the product of efforts to provide common records was judged strictly in terms of the "selfish" or particularistic needs of some medical and other related research groups; the notion has never gained really wide support.

Many other examples might be given of special needs. There is the case, for example, of the administrator of a chemical research facility who still thinks of himself as a "bench chemist," because that image is closely linked to his perceived status as a scientist. In point of fact, although his lab coat may be changed every week, and is hung in the corner of his office, it has not been worn for twenty years at a bench. Such persons, like deans of medicine, may support strongly any proposal for a general information system for chemistry, or medicine, or metals technology; but their staffs often will not use them.

It must also be remembered that there are probably relatively few people who really care about using large amounts of information. Most people tend to deal with the world around them on a rather superficial basis. The journals they receive are very often articles of "furniture," used to decorate offices and study dens; neither the content nor the table of contents may be used. If one is interested in "engaging" these people, then one must make allowance for a superficial type of scanning activity.

There are ways to find out what users need what information relatively cheaply; such is the approach ERIC is using. The technique is to extent the utilization of interpretive summaries of the literature of a field. This approach may work up to some as yet unspecified level. We know, for example, that there is one abstract journal for about every 300 published journals. At present, with more than 300 abstract journals now in existence, we have two abstract-of-abstracts journals. There comes a point at which the distance between the user and the literature he searches becomes great enough that he will lose the sense of relevance of the literature to what he is doing in his own work. But at least we will find this out without first creating a very elaborate, and very expensive, fully developed retrieval system. It should always be kept in mind, too, that no information system has ever changed the information-using behavior of the people who use it; in terms of our knowledge of behavior, we have evidence about the possibility of changing the information-acquiring habits of people, but little or nothing has been learned about how they use it.

In designing general purpose information systems, we tend to take as a criterion of success whether or not they are used by the populations they are designed for. If one designs a system for large enough "cuts" of the population, of course, one can hardly miss having the system used (such is ERIC's virtue). With a population of specialized users, however, such as those interested in sensory research, it is possible to miss them all.

THE QUESTION OF EVALUATION

Another way to consider the relevancy of the user profile, the match to system characteristics, and the value of the system to the user, is to investigate the effectiveness of the man/machine interrelationship in practice. Even with this shift in emphasis, we are still dealing with many of the problems already discussed. How can we evaluate the match between the need to know and the satisfaction of that need?

One can make a market survey, as we have indicated, and develop a profile of the user and his range of interests. This should be done before the system is built, not after, as we noted above. In fact, no one has approached a system design in this way to date.

One can also identify the leaders in a field, and try to respond to their needs for information, at least as a "first pass." The adequacy of this approach will be revealed only with use. Certainly it seems plain that the system should be one actually used by the designers and builders of the system; if it is not useful to them, it is not likely to be used by anyone else. One of the difficulties here is that a certain amount of money and time and trouble must be devoted to a system before it will "go critical," that is, will have the capability for handling information minimally effective in supplying data when needed. So far as we can see now, this implies, first of all, bibliographic control over the literature of a field. The task of probing research data is essentially a task of finding bibliographic information; these are tasks of screening and refereeing. Ideally, the system will also be educative, in the sense that it will question the user, and respond in terms of narrowing a file search request. Thus, no research system can be divorced from a bibliographic system. The system will become useful to a nonspecialist when it includes syntheses of the literature, and other information, also available through the "educative" mode of the user/machine interaction; the major limitation on this structure is the amount of money available to create it.

But to the extent that the system can accommodate itself to the needs of the nonscholar, it will become useful to a wider population of teachers, of administrators, of students, of neophytes.

Unfortunately no system has as yet seriously considered the difficulties of evaluation of user response to the operation of a system over time. One is tempted to think in terms of the cost/effectiveness criteria of systems research, developed in the military. Unfortunately, the criteria used in military operations research are simplistic and concrete--hardly suited to judging the effectiveness of information-using behaviors in man/machine systems. The crucial variable for evaluating system effectiveness is the "satisfaction" of the user (and we are deliberately allowing the imprecision of this term); but "satisfaction" is not so amenable to counting and analysis as is speed of recovery, or the number of "false drops." It appears we shall have to develop a desideratum space, in which we will consider not only what the user is looking for, but also the kind of "thing" he is after in addressing the system.

What seems called for is a constant and iterative process, which measures the user's satisfaction within the desideratum space, and uses the result to modify the data base of the system, and perhaps also the operation of the system itself. This is half the story, only; the other half is of course the amount of money that the user (or whomever is paying for the system) is willing to spend to achieve any given level of "satisfaction." Let it be admitted that user satisfaction may be achieved in a small and inexpensive information system, just as likely as in a large, complex, and costly system--at a given point in the desideratum space. There has been a number of cases in which an elaborate information system has failed because users were not willing to pay the costs of operating a system which met the minimum demands of their search strategies; and the systems which incorporate a well-understood pricing factor for satisfying search requests appear to have the best chance of survival and growth. This is in fact the crux of the argument between general vs. special purpose systems: Population A may be unwilling to pay the increment in cost for satisfaction of its needs and those of Population B; the response may be, "Let Population B go develop its own system." And this may be the most economical solution. So far as the introduction of technology into information retrieval is concerned, one might well take the posture that he will only be dragged kicking and screaming into the 21st century--and only when the increased complexity and cost result in a significant increment in the "satisfaction" of the user with the information product he receives from the system.

THE ROLE OF EXPERIENCE

The evaluation of information systems, like the evaluation of sensory aids, presupposes a period of training of the neophyte in the use of the system; this is so, at least, if one wishes to evaluate the system and not merely the reaction to the system. In the instructional materials centers of the ERIC system, experience has shown that exposing practitioners to the efficiency of the information system in satisfying their need to know, and teaching them how to use the system, in effect builds up a population of users where none had existed before. Subject to the limitations of cost, the satisfaction of sophisticated search request strategies, and the definition of the perceived needs of users (all of which we have discussed as caveats in foregoing sections), this approach seems to work well. Best of all, it would appear, is the interactive mode of file searching, now an inherent part of such systems as that of the Harvard Vision Information Center (one of the NIND satellite systems in the MEDLARS linkage). This system approach permits the fastest possible utilization of user experience in modifying the system, in narrowing search requests in real time, and in providing record-keeping capabilities to aid systems designers. Its most unique characteristic is that the user is instructed how to best utilize the stored data in the system by means of a stepped program; he learns to use the system at the same time he is addressing search requests to the system.

The crucial role of the experience of users with the information systems under evaluation is that one cannot merely segment the whole population into parts; there is no clean set of segmentation lines. One may have many alternative ways of categorizing users, and we have no way of knowing what segmentation puts together those characteristics most useful in designing systems. It may be, as suggested above, that the most efficient solution will be to design systems satisfying the needs of subpopulations.

This conclusion comes as a shock to many system designers, who believed in the ability of large-scale general purpose systems to satisfy all user needs. But there is no inherent reason to expect that communication will be better between researchers and practitioners in one field than in another. The value of using all the data we can generate about actual use of systems--as in interactive modes, and with continual and iterative efforts to fix the values of the desideratum space for the "satisfaction" of users--is that we may thereby realize the hope that we can design systems uniquely well suited to the information needs of a given population of users, whether specialists or neophytes, researchers or practitioners, experts or students.

THE IRIS DOCUMENT CENTER

The IRIS Document Center uses a simple and inexpensive document system, amenable to mechanization, to exert bibliographic control over a specialized literature in sensory research and sensory aids research and development. The system builders did not really know who would use the system they designed; it was hoped it would be useful to researchers, like themselves, who were not specialists in many of the areas covered, and who sought a short cut through the many disparate fields of research in which work in this area was embedded. But there was no effort to design a system oriented to only one type of user, since it was hoped that the time-saving would be of equal value to those persons who were not researchers, but who were interested in what researchers were doing in this field. Among the persons who have used the system so far--and it is not yet fully operational--are teachers, students, researchers, and administrators. The value of the system seems to be greatest for nonspecialists in the area of sensory research, i.e., researchers and others who have not had a commitment to this field. It is therefore most useful to the "nonestablishment."

Because the system is still under development, it is still possible, maybe desirable, to detach certain areas of concern from its data base, and depend on other specialized information centers in a cooperative relationship with IRIS to deal with certain aspects of the specialized literature. Thus, IRIS might depend on the Western Michigan Instructional Materials Center for educational research and practice related to the blind; and on the Harvard Vision Information Center for medical data on blindness.

The approach used in the system is simple: to provide information about research in blindness to those who wish to know something about it. We cannot say who should use the system, nor what the criteria should be for using the system. And at present, there is no provision for evaluating user satisfaction.

CONCLUSION

We started by asking what guidelines the "big" system operators could give to specialized information center system builders. This review does convey most of the concerns we expressed in answering that question. As an exercise, then, we might ask, how would we paraphrase some "practical" advice to a designer concerned with a population of users with a special interest? Among the matters we might wish to bring to his attention would be the following:

1. Identify the leaders in your field and enlist their efforts in stating their needs for information storage and retrieval.

2. Design your system to take account of these needs, leaving room for modification of the system through use, and reducing the turn-around time from inquiry to system response to the minimum; this may require you to consider seriously the provision of an interactive mode of inquiry.

3. Now expose practitioners in your field (i.e., those who will use the information generated by the leaders) to the system, instructing them in its use, and modifying the system to take account of the satisfaction, or lack of it, that the practitioner experiences

in seeking information. Some guides to modification in use may be found in the experience of similarly designed systems, such as the Instructional Materials Center in Texas, and the Dartmouth College Time Sharing System.

4. Make a charge for each increment in user satisfaction which requires modification; when these become too expensive, consider the creation of alternative systems to satisfy the information needs of subpopulations. (Sensory research may indeed be specialized enough to require a separate information system.)

5. Remember that we have knowledge developed through behavioral research of the information *acquiring* behavior of persons who use information systems, but virtually no comparable knowledge about the information *using* behavior of people. You will have the difficult job of designing a system for which criteria of user satisfaction, and data on how users actually put information to use, remain to be developed.

6. If these problems and lack of data decrease your motivation to design an elaborate system, consider the possibilities of a least cost alternative, that of exerting bibliographic control with rapid access, as a first step toward making possible the conditions under which the missing data may be generated. Remember that you will have to add to the system a set of experiments to help generate the data, and you will meet considerable resistance and not a few problems in the conduct of the needed research. You will obviously want to depend on independent efforts of information science specialists to bring this needed data into existence.

The above is but a very simplified version of the concerns developed at length in several of the papers in this volume-- papers which will repay careful reading. If a slightly negative air hangs around the remarks made here, it may be attributed to two factors. One is the pressing, and constantly increasing, need for fast answers to immediate problems; this is responsible for the anxiety we feel about developing adequate information systems quickly. The other is the very common tendency to assume that a mechanized solution to our needs for information must succeed, that we can produce a solution to the problem ex machina, particularly ex computer; and that information specialists have the knowledge and the resources to do so right at the moment. It is useful to remind ourselves that research in these areas, which is made difficult also because it requires the close cooperation and understanding of researchers and practitioners in the field of information sciences, is a cooperative enterprise; moreover, like other areas of research, it may raise more problems than it solves, it is slow, it is relatively expensive, and it offers only the guarantee of greater understanding of the problems it attacks, not necessarily their solution.

The difficulty of the problems we have dealt with need not obscure the fact that the organizers of this conference, like many participants, felt stimulated by the exercise of making our problems explicit, instead of struggling with apparently indominatable monsters in private. And we were hopeful and encouraged by the best examples of the systems described that bibliographic control can be exerted over a body of literature; that we can design a system to interact in real time with the user; that we can build instruction in the use of a system into the activity of searching for information; and that we can at least envisage the possibilities for finding out some crucial characteristics of information *using* behavior, and using that knowledge to increase both our measurement of, and ability to create greater, user satisfaction with our systems.

If we look back to the beginning of the "information explosion" which gave rise to many of the concerns we have dealt with, and realize that only a little more than a decade separates us from it, we have good reason to be pleased with our ability to pose better questions than we have in the past. This is, of course, the beginning of wisdom. We look forward to our next opportunity to discuss our growth in wisdom, and hope that we can show a still greater increment in our satisfaction with what we have learned.

PARTICIPANTS

Dr. Jordan J. Baruch
Former President
EDUCOM
100 Charles River Plaza
Boston, Massachusetts 02114

Leslie L. Clark
Director
International Research Information Service
American Foundation for the Blind
15 West 16th Street
New York, New York 10011

George J. Howick
Executive Vice President
IRM (USA) Incorporated
The Plaza, 2 West 59th Street
New York, New York 10019

Dr. Kenneth R. Ingham
Research Associate
Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

Allen Kent
Director
Knowledge Availability Systems Center
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

Dr. Joseph Leiter
Associate Director
National Library of Medicine
8600 Rockville Pike
Bethesda, Maryland 20014

James W. Linsner
Former Supervisor
IRIS Document Center
American Foundation for the Blind
15 West 16th Street
New York, New York 10011

Dr. Claude H. Marks
Associate Director
Instructional Materials Center
University of Texas
Austin, Texas 78712

Mr. Harvey Marron
Director
ERIC Information Centers
Bureau of Information, Technology, and Dissemination
United States Office of Education
Washington, District of Columbia 20202

Dr. Paul H. Masoner
Dean
School of Education
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

Mr. Edison Montgomery
Vice Chancellor
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

Dr. Ralph Peabody
Assistant Professor
Department of Special Education and Rehabilitation
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

Dr. Samuel T. Price
Department of Special Education
Illinois State University
Normal, Illinois 61761

Dr. Robert Reinecke
Howe Laboratory of Ophthalmology
Harvard University Medical School
243 Charles Street
Boston, Massachusetts 02114

Dr. Godfrey D. Stevens
Professor
Special Education and Rehabilitation
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

Dr. Ted Ward
Director
Learning Systems Institute
The Michigan State University
343-B Erickson Hall
East Lansing, Michigan 48823